MADURAI KAMARAJ UNIVERSITY

(University with Potential for Excellence)

Madurai – 625 021.

DIRECTORATE OF DISTANCE EDUCATION

M.Sc., Botany

Paper – IV

Cell Biology Genetics, Plant Breeding and Evolution

www.mkudde.org

6046

PBOT04

Printed At Gokul Agencies

Copies – 500 Fresh Print -2014

PAPER IV: CELL BIOLOGY GENETICS, PLANT BREEDING AND EVOLUTION

Structural Organization: Prokaryotic and Eukaryotic cell. Cell wall- Primary and secondary Structure, model and function. Plasma membrane-channels, Pumps and Receptors; plasmodesmata. Cell organelles –structure and functions; Ultra structure and semi autonomous nature of chloroplast and mitochondria.

Structure and function of Organelles- Nucleus, ER, Golgi complex, ribosome, Lysosome, Chromosome- Lamp brush, polytene. Microscopy – light microscope, compound microscope, Electron microscope, dark field microscope, phase contrast Microscope, Cell cycle and cell division.

Identification of DNA as genetic material – Watson and crick model of DNA, RNA, si RNA & ri RNA. DNA replication – Transcription- Prokaryotic RNA polymerase- Types, initiation, elongation, Termination- genetic code and Protein synthesis- Post translational modifications.

Operon Concepts – Mechanism of gene regulation. Lac, Ara & trp regulation in prokaryotes, Gene regulation in Eukaryotes. Cytogenetics of polyploidy & aneuploids. Mutations – spontaneous and induced. Molecular basis of mutation; DNA demage and repair mechanism.

Plant breeding: Introduction and scope, pure line selection- Mass selection pedigree method, Bulk Method, Back cross method and clonal selection and hybridization.

Origin of life- theories of evolution –Lamarckism, Darwinism, Mutation theory, modern synthetic theory – types of evolution and evolution in action –speciation- types and mechanism of speciation –adaptation, polymorphism and co-evolution

Text Books:

1. De Robertis, E.D.P. and robertis, E.M.C.2004. *Cell Biology*. B.I. Waverly Pvt., Ltd., New Delhi.
2. Freifelder, D.2000. *Molecular Biology*. II Edn., Narosa publishing House New Delhi.
3. Gupta, M.I. and Jangir, M.L.2003. *Cell Biology*: Fundamentals and applications. Agrobios, Jodpur.
4. Singh, B.D.2001. *Principles of Plant Breeding*. Kalyani Publishers, Ludhiana

Reference Books:

1. Albers, B., Bray, D., Lewis, J., Raff, Roberts, K. and Watson, J.D.1999.*Molecular Biology of the cell* (3rd Edition).Garland publishing, Inc., New York.
2. De, D.N.2000. *Plant Cell vacuoles*: An Introduction. CSIRO publication, Collingwood, Australia.

3. Lewin, B.2000.Genes VII. Oxford University Press, New York.

4. Rost, T.et al.1998. *Plant Biology*. Wads worth Publishing co., California, USA,

5. Wolfe, S.L.1993.Molecluar and Cellular Biology. Wadsworth Publishing Co., California, USA

CONTENTS

UNIT	CHAPTER	PAGE.NO
1	Structural Organization	1 - 19
2	Structures and Function of Organelles	20 - 39
3	Identification Of DNA As Genetic Material	40 - 89
4	Plant Breeding	90 - 126
5	Origin of Life	127 - 156

UNIT – 1 STRUCTURAL ORGANIZATION

The Origin and Evolution of Cells

Cells are divided into two main classes, initially defined by whether they contain a nucleus. Prokaryotic cells (bacteria) lack a nuclear envelope; eukaryotic cells have a nucleus in which the genetic material is separated from the cytoplasm. Prokaryotic cells are generally smaller and simpler than eukaryotic cells; in addition to the absence of a nucleus, their genomes are less complex and they do not contain cytoplasmic organelles or a cytoskeleton (Table 1). In spite of these differences, the same basic molecular mechanisms govern the lives of both prokaryotes and eukaryotes, indicating that all present-day cells are descended from a single primordial ancestor.

Table – 1 Prokaryotic and Eukaryotic Cell

Characteristic	**Prokaryote**	**Eukaryote**
Nucleus	Absent	Present
Diameter of a typical cell	1 µm	10 – 100 µm
Cytoskeleton	Absent	Present
Cytoplasmic organelles	Absent	Present
DNA content (base pairs)	1×10^6 to 5×10^6	1.5×10^7 to 5×10^9
Chromosomes	Single circular DNA molecule	Multiple linear DNA molecule

Present-Day Prokaryotes

Present-day prokaryotes, which include all the various types of bacteria, are divided into two groups-the archaebacteria and the eubacteria which diverged early in evolution. Some archaebacteria live in extreme environments, which are

unusual today but may have been prevalent in primitive Earth. For example, thermoacidophiles live in hot sulfur springs with temperatures as high as 80°C and pH values as low as 2. The eubacteria includethe common forms of present-day bacteria-a large group of organisms that live in a wide range of environments, including soil, water, and other organisms (e.g., human pathogens).

Most bacterial cells are spherical, rod-shaped, or spiral, with diameters of to 10 pm. Their DNA contents range from about 0.6 million to 5 million base pairs, an amount sufficient to encode about 5000 different proteins. The largest and most complex prokaryotes are the cyanobacteria, bacteria in which photosynthesis evolved.

The structure of a typical prokaryotic cell is illustrated by Escherichia coli (E. coli), a common inhabitant of the human intestinal tract. The cell is rod-shaped, about 1 pm in diameter and about 2 pm long. Like most other prokaryotes, E. coli is surrounded by a rigid cell wall composed of polysaccharides and peptides. Beneath the cell wall is the plasma membrane, which is a bilayer of phospholipids and associated proteins. Whereas the cell wall is porous and readily penetrated by a variety of molecules, the plasma membrane provides the functional separation between the inside of the cell and its external environment. The DNA of E. coli is a single circular molecule in the nucleoid, which, in contrast to the nucleus of eukaryotes, is not surrounded by a membrane separating it from the cytoplasm. The cytoplasm contains approximately 30,000 ribosomes (the sites of protein synthesis), which account for its granular appearance.

Eukaryotic Cells

Like prokaryotic cells, all eukaryotic cells are surrounded by a plasmamembrane and contain ribosomes. However, eukaryotic cell s are much more complex and contain a nucleus, a variety of cytoplasmic organelles, and a cytoskeleton. The largest and most prominent organelle of eukaryotic cells is the nucleus, with a diameter of approximately 5 pm. The nucleus contains the genetic information of the cell, which in eukaryotes is organized as linear rather than circular DNA molecules. The nucleus is the site of DNA replication and of RNA synthesis; the translation of RNA into proteins takes place on ribosomes in the

cytoplasm. In addition to a nucleus, eukaryotic cells contain a variety of membrane enclosed organelles within their cytoplasm. These organelles provide compartments in which different metabolic activities are localized. Eukaryotic cells are generally much larger than prokaryotic cells, frequently having a cell volume at least a thousand fold greater. The compartmentalization provided by cytoplasmic organelles is what allows eukaryotic cells to function efficiently. Two of these organelles, mitochondria and chloroplasts, play critical roles in energy metabolism. Mitochondria, which are found in almost all eukaryotic cells, are the sites of oxidative metabolism and are thus responsible for generating most of the ATP derived from the breakdown of organic molecules. Chloroplasts are the sites of photosynthesis and are found only in the cells of plants and green algae. Lysosomes and peroxisomes also provide specialized metabolic compartments for the digestion of macromolecules and for various oxidative reactions, respectively. In addition, most plant cells contain large vacuoles that perform a variety of functions, including the digestion of macromolecules and the storage of both waste products and nutrients. Because of the size and complexity of eukaryotic cells, the transport of proteins to their correct destinations within the cell is a formidable task. Two cytoplasmic organelles, the endoplasmic reticulum and the Golgi apparatus, are specifically devoted to the sorting and transport of proteins destined for secretion, incorporation into the plasma membrane, and incorporation into lysosomes. The endoplasmic reticulum is an extensive network of intracellular membranes, extending from the nuclear membrane throughout the cytoplasm. It functions not only in the processing and transport of proteins, but also in the synthesis of lipids. From the endoplasmic reticulum, proteins are transported within small membrane vesicles to the Golgi apparatus, where they are further processed and sorted for transport to their final destinations. In addition to this role in protein transport, the Golgi apparatus serves as a site of lipid synthesis and (in plant cells) as the site of synthesis of some of the polysaccharides that compose the cell wall. Eukaryotic cells have another level of internal organization: the cytoskeleton, a network of protein filaments extending

throughout the cytoplasm. The cytoskeleton provides the structural framework of the cell, determining cell shape and the general organization of the cytoplasm.

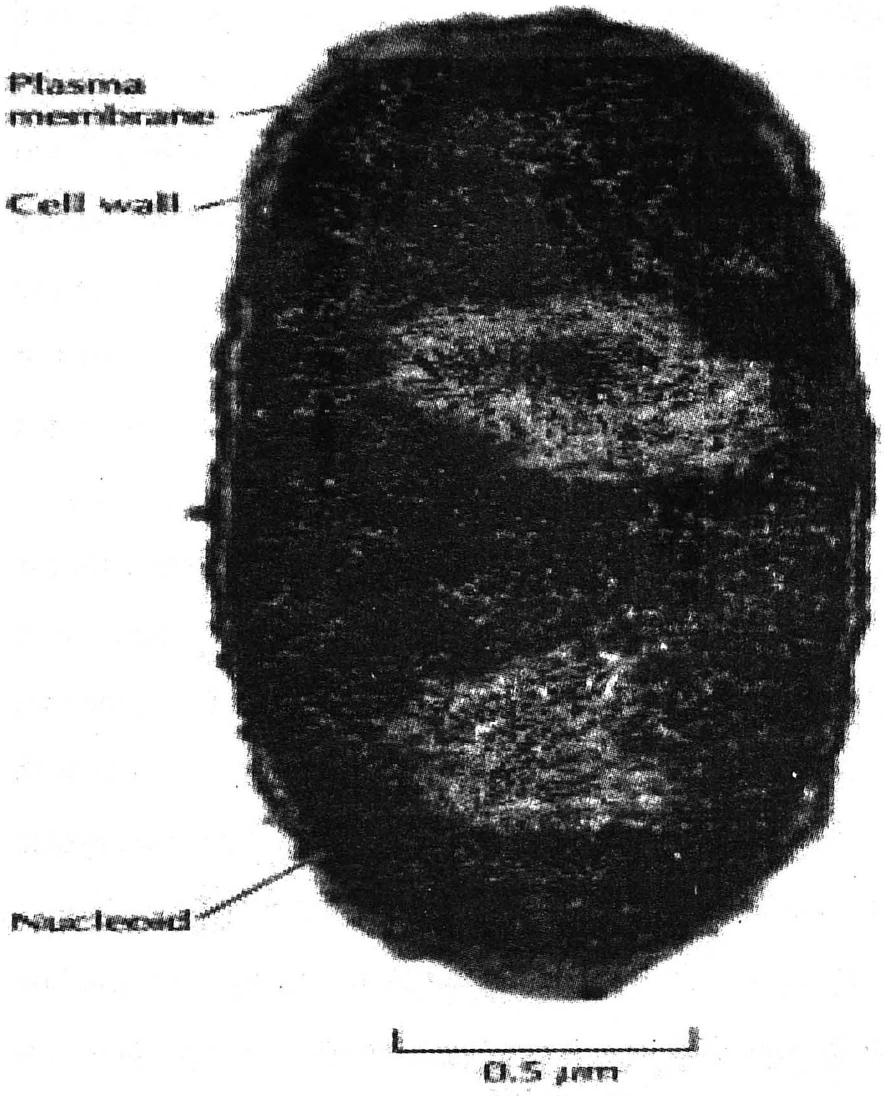


Figure-1 Electron micrograph of E. coli the cell is surrounded by a cell wall, beneath which is the plasma membrane. DNA is located in the nucleoid.

Structure of the Plasma Membrane

Like all other cellular membranes, the plasma membrane consists of both lipids and proteins. The fundamental structure of the membrane is the phospholipid bilayer, which forms a stable barrier between two aqueous compartments. In the case of the plasma membrane, these compartments are the inside and the outside of the cell. Proteins embedded within the phospholipid bilayer carry out the specific functions of the plasma membrane, including selective transport of molecules and cell-cell recognition.

The Phospholipid Bilayer

The plasma membrane is the most thoroughly studied of all cell membranes, and it is largely through investigations of the plasma membrane that our current concepts of membrane structure have evolved. The plasma membranes of mammalian red blood cells (erythrocytes) have been particularly useful as a model for studies of membrane structure. Mammalian red blood cells do not contain nuclei or internal membranes, so they represent a source from which pure plasma membranes can be easily isolated for biochemical analysis. Indeed, studies of the red blood cell plasma membrane provided the first evidence that biological membranes consist of lipid bilayers. ln 1925, two Dutch scientists (Edwin Gorter and F. Grendel) extracted the membrane lipids from a known number of red blood cells corresponding to a known surface area of plasma membrane. They then determined the surface area occupied by a monolayer of the extracted lipid spread out at an air-water interface. The surface area of the lipid monolayer turned out to be twice that occupied by the erythrocyte plasma membranes, leading to the conclusion that the membranes consisted of lipid bilayers rather than monolayers. The bilayer structure of the erythrocyte plasma membrane is clearly evident in high magnification electron micrographs. The plasma membrane appears as two dense lines separated by an intervening space a morphology frequently referred to as a "railroad track" appearance. This image results from the binding of the electron-dense heavy metals used as stains in transmission electron microscopy (see Chapter 1) to the polar head groups of the phospholipids, which therefore appear as dark lines. These dense lines are separated by the lightly stained interior portion of the membrane, which contains the hydrophobic fatty acid chains.

The plasma membranes of animal cells contain four major phospholipids (phosphatidylcholine, phosphatidyleth anolamine, phosphatidyl serine, and sphingomyelin), which together account for more than half of the lipid in most membranes. These phospholipids are asymmetrically distributed between the two halves of the membrane bilayer. The outer leaflet of the plasma membrane consists mainly of phosphatidylcholine and sphingomyelin, whereas phosphatidylethanolamine and phosphatidylserine are the predominant

phospholipids of the inner leaflet. The head groups of both phosphatidylserine and phosphatidylinositol are negatively charged, so their predominance in the inner Leaflet results in a net negative charge on the cytosolic face of the plasma membrane. A fifth phospholipid, phosphatidylinositol, is also localized to the inner half of the plasma membrane. Although phosphatidylinositol is a quantitatively minor membrane component, it plays an important role in cell signalling. In addition to the phospholipids, the plasma membranes of animal cells contain glycolipids and cholesterol. The glycolipids are found exclusively in the outer leaflet of the plasma membrane, with their carbohydrate portions exposed on the cell surface. They are relatively minor membrane components, constituting only about 2% of the lipids of most plasma membranes. Cholesterol, on the other hand, is a major membrane constituent of animal cells, being present in about the same molar amounts as the phospholipids. Two general features of phospholipid bilayers are critical to membrane function. First, the structure of phospholipids is responsible for the basic function of membranes as barriers between two aqueous compartments. Because the interior of the phospholipid bilayer is occupied by hydrophobic fatty acid chains, the membrane is impermeable to water-soluble molecules, including ions and most biological molecules. Second, bilayers of the naturally occurring phospholipids are viscous fluids, not solids. The fatty acids of most natural phospholipids have one or more double bonds, which introduce kinks into the hydrocarbon chains and make them difficult to pack together. The long hydrocarbon chains of the fatty acids therefore move freely in the interior of the membrane, so the membrane itself is soft and flexible. In addition, both phospholipids and proteins are free to diffuse laterally within the membrane-a property that is critical for many membrane functions. Because of its rigid ring structure, cholesterol plays a distinct role in membrane structure. Cholesterol will not form a membrane by itself but inserts into a bilayer of phospholipids with its polar hydroxyl group close to the phospholipid head groups. Depending on the temperature, cholesterol has distinct effects on membrane fluidity. At high temperatures, cholesterol interferes with the movement of the phospholipid fatty acid chains, making the outer part of the membrane less fluid and reducing its

permeability to small molecules. At low temperatures, however, cholesterol has the opposite effect: By interfering with interactions between fatty acid chains, cholesterol prevents membranes from freezing and maintains membrane fluidity.

Although cholesterol is not present in bacteria, it is an essential component of animal cell plasma membranes. Plant cells also lack cholesterol, but they contain related compounds (sterols) that fulfil a similar function. Rather than diffusing freely in the plasma membrane, cholesterol and the sphingolipids (sphingomyeli n and glycolipids) form discrete membrane domains known as lipid rafts. These clusters of sphingolipids and cholesterol move laterally within the plasma membrane and associate with specific membrane proteins. Although the functions of lipid rafts remain to be fully understood, they play important roles in processes such as cell movement and the uptake of extracellular molecules by endocytosis as well as in cell signalling.

Membrane Proteins

While lipids are the fundamental structural elements of membranes, proteins are responsible for carrying out specific membrane functions. Most plasma membranes consist of approximately 50% lipid and 50% protein by weight, with the carbohydrate portions of glycolipids and glycoproteins constituting 5 to 10% of the membrane mass. Since proteins are much larger than lipids, this percentage corresponds to about one protein molecule per every 50 to 100 molecules of lipid. In 1972 Jonathan Singer and Garth Nicolson proposed the fluid mosaic model of membrane structure, which is now generally accepted as the basic paradigm for the organization of all biological membranes. In this model, membranes are viewed as two-dimensional fluids in which proteins are inserted into lipid bilayers (Figure 3).

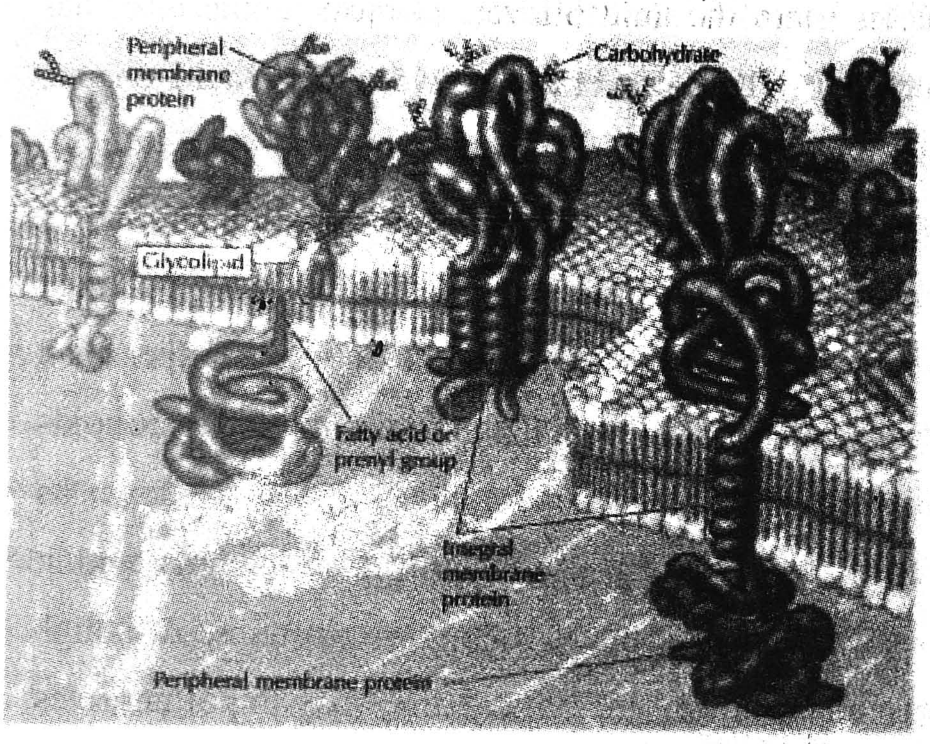


Figure-3

Fluid mosaic model of the plasma membrane integral membrane proteins are inserted into the lipid bilayer, whereas peripheral proteins are bound to the membrane indirectly by protein-protein interactions.Most integral membrane proteins are transmembrane proteins with portions exposed on both sides of the lipid bilayer. The extracellular portions of these proteins are usually glycosylated, as are the peripheral membrane proteins bound to the external face of the membrane.

Singer and Nicolson distinguished two classes of membrane-associated proteins, which they called peripheral and integral membrane proteins. Peripheral membrane proteins were operationally defined as proteins that dissociate from the membrane following treatments with polar reagents, such as solutions of extreme pH or high salt concentration that do not disrupt the phospholipid bilayer. Once dissociated from the membrane, peripheral membrane proteins are soluble in aqueous buffers. These proteins are not inserted into the hydrophobic interior of the lipid bilayer. Instead, they are indirectly associated with membranes through protein-protein interactions. These interactions frequently involve ionic bonds, which are disrupted by extreme pH or high salt. In contrast to the peripheral membrane proteins, integral membrane proteins can be released only by treatments that disrupt the phospholipid bilayer. Portions of these integral

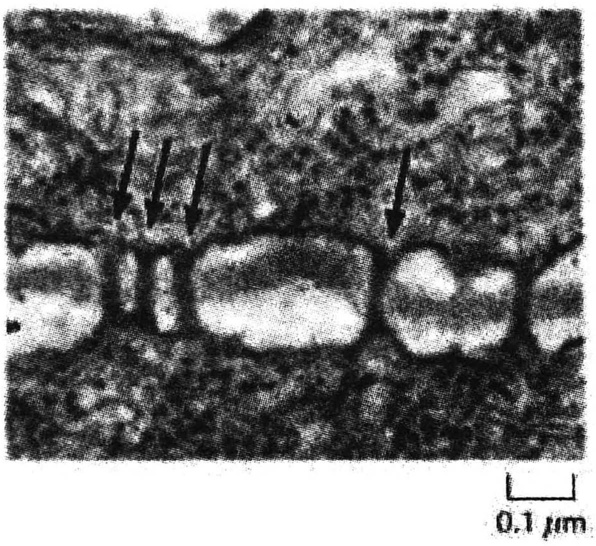
membrane proteins are inserted into the lipid bilayer, so they can be dissociated only by reagents that disrupt hydrophobic interactions. The most commonly used reagents for solubilization of integral membrane proteins are detergents, which are small amphipathic molecules containing both hydrophobic and hydrophilic groups. The hydrophobic portions of detergents displace the membrane lipids and bind to the hydrophobic portions of integral membrane proteins. Because the other end of the detergent molecule is hydrophilic, the detergent protein complexes are soluble in aqueous solutions. Many integral proteins are transmembrane proteins, which span the lipid bilayer with portions exposed on both sides of the membrane. These proteins can be visualized in electron micrographs of plasma membranes prepared by the freeze fracture technique (see Figure 1.36). In these specimens, the membrane is split and separated into its two leaflets. Transmembrane proteins are then apparent as particles on the internal faces of the membrane. The membrane-spanning portions of transmembrane proteins are usually a helices of 20 to 25 hydrophobic amino acids that are inserted into the membrane of the endoplasmic reticulum during synthesis of the polypeptide chain. These proteins are then transported in membrane vesicles from the endoplasmic reticulum to the Golgi apparatus and from there to the plasma membrane. Carbohydrate groups are added to the polypeptide chains in both the endoplasmic reticulum and Golgi apparatus, so most transmembrane proteins of the plasma membrane are glycoproteins with their oligosaccharides exposed on the surface of the cell. Studies of red blood cells have provided good examples of both peripheral and integral proteins associated with the plasma membrane. The membranes of human erythrocytes contain about a dozen major proteins, which were originally identified by gel electrophoresis of membrane preparations. Most of these are peripheral membrane proteins that have been identified as components of the cortical cytoskeleton, which underlies the plasma membrane and determines cell shape. For example, the most abundant peripheral membrane protein of red blood cells is spectrin, which is the major cytoskeletal protein of erythrocytes. Other peripheral membrane proteins of red blood cells include actin, ankyrin, and band 4.1. Ankyrin serves as the principal link between the plasma membrane and the

cytoskeleton by binding to both spectrin and the integral membrane protein band 3. An additional link between the membrane and the cytoskeleton is provided by band 4. 1, which binds to the junctions of spectrin and actin, as well as to glycophorin (the other major integral membrane protein of erythrocytes). The two major integral membrane proteins of red blood cells glycophorin and band 3- provide well sudied examples of transmembrane protein structure. Glycophorin is a small glycoprotein of 131amino acids, with a molecular weight of about 30,000, half of which is protein and half carbohydrate. Glycophorin crosses the membrane with a single membrane-spanning a helix of 23 amino acids, with its glycosylated amino terminal portion exposed on the cell surface. Although glycophorin was one of the first transmembrane proteins to be characterized, its precise function remains unknown. In contrast, the function of the other major transmembrane protein of red blood cells is well understood. This protein, originally known as band 3, is the anion transporter responsible for the passage of bicarbonate (HC03) and chloride (CJ-) ions across the red blood cell membrane. The band 3 polypeptide chain is 929 amino acids and is thought to have 14 membrane-spanning a-helical regions. Within the membrane, dimers of band 3 form globular structures containing internal channels through which ions are able to travel across the lipid bilayer.

Plasmodesmata

Adhesion between plant cells is mediated by their cell walls rather than by transmembrane proteins. In particular, a specialized pectin-rich region of the cell wall called the middle lamella acts as a glue to hold adjacent cells together. Because of the rigidity of plant cell walls, stable associations between plant cells do not require the formation of cytoskeletallinks such as those provided by the desmosomes and adherens junctions of animal cells. However, adjacent plant cells communicate with each other through cytoplasmic connections called plasmodesmata (singular, plasmodesma). Although distinct in structure, plasmodesmata function analogously to gap junctions as a means of direct communication between adjacent cells in tissues. Plasmodesmata form from

incomplete separation of daughter cells following plant cell mitosis. At each plasmodesma, the plasma membrane of one cell is continuous with that of its neighbor, creating a channel between the two cytosols. An extension of the smooth endoplasmic reticulum passes through the pore, leaving a ring of surrounding cytoplasm through which ions and small molecules are able to pass freely between the cells. Plasmodesmata are dynamic structures that can open or close in response to appropriate stimuli, permitting the regulated passage of macromolecules between adjacent cells. In addition, there is evidence that proteins and lipids can be targeted to plasmodesmata in response to specific signals. Plasmodesmata may thus play a key role in plant development by controlling the trafficking of regulatory molecules, such as transcription factors or RNAs, between cells.



Chloroplast:

Chloroplasts are one of the many different types of organelles in the plant cell. They are considered to have originated from cyanobacteria through endosymbiosis when a eukaryotic cell engulfed a photosynthesizing cyanobacterium which remained and became a permanent resident in the cell. This was first suggested by Mereschkowsky in 1905 after an observation by Schimper in 1883 that chloroplasts closely resemble cyanobacteria. Chloroplasts are similar to mitochondria in that they both originate from an endosymbiotic event, but chloroplasts are found only in plants and protista.

In some groups of mixotrophic protists such as the dinoflagellates, chloroplasts are separated from a captured alga or diatom and used temporarily. These to

chloroplasts may only have a lifetime of a few days and are then replaced. In land plants, chloroplasts are generally lens-shaped, 5–8 μm in diameter and 1–3 μm thick. They are larger than mitochondria due to the fact that their internal membranes are not folded up into cristae. The chloroplast is contained by an envelope that consists of an inner and an outer membrane. Between these two layers is the intermembrane space.

A typical chlorenchyma cell of a land plant contains about 10 to 100 chloroplasts. In the cells of many alga there is only one chloroplast (for example in *Chlorella*, it fills much of the cell and is bell-shaped).

Recent studies have shown that chloroplasts can be interconnected by tubular bridges called stromules, formed as extensions of their outer membranes.[43][44] Chloroplasts appear to be able to exchange proteins via stromules, and thus function as a network.

Stroma

The fluid within the chloroplast is called the stroma, corresponding to the cytosol of the original cyanobacterium. Circular nucleoids of chloroplast DNA and chloroplast ribosomes can be found floating around in it. The Calvin cycle, which fixes CO_2 into sugar takes place in the stroma.

Thylakoids

Within the stroma are stacks of thylakoids, which are the site of the light reactions of photosynthesis. The thylakoids are arranged in stacks called grana (singular: granum). A thylakoid has a flattened disk shape. Inside it is an cavity called the thylakoid space or thylakoid lumen.

In the transmission electron microscope, thylakoid membranes appear as alternating light-and-dark bands, each 0.01 μm thick. Embedded in the thylakoid membrane are antenna complexes, each of which consists of the light-absorbing pigments, including chlorophyll and carotenoids, as well as proteins that bind the pigments. This complex both increases the surface area for light capture, and allows capture of photons with a wider range of wavelengths. The energy of the incident photons is absorbed by the pigments and funneled to the reaction center

of this complex through resonance energy transfer. Two chlorophyll molecules are then ionized, producing an excited electron, which then passes onto the photosynthetic reaction center.

The number of thylakoids and the total thylakoid area of a chloroplast is influenced by light exposure. Shaded chloroplasts contain larger and more grana with more thylakoid membrane area than chloroplasts exposed to bright light, which have smaller and fewer grana and less thylakoid area. Thylakoid extent can change within minutes of light exposure or removal.

Pyrenoids

The chloroplasts of some hornworts and algae contain structures called pyrenoids. They are not found in higher plants. Pyrenoids are roughly spherical and highly refractive bodies which are a site of starch accumulation in plants that contain them. They consist of an matrix opaque to electrons, surrounded by two hemispherical starch plates. The starch is accumulated as the pyrenoids mature. In algae with carbon concentrating mechanisms, the enzyme rubisco is found in the pyrenoids. Starch can also accumulate around the pyrenoids when CO_2 is scarce. Pyrenoids can divide to form new pyrenoids, or be produced "de novo".

Peripheral reticulum

Some chloroplasts contain a structure called the chloroplast peripheral reticulum. It is often found in the chloroplasts of C_4 plants, though it's also been found in some C_3 angiosperms, and even some gymnosperms. The chloroplast peripheral reticulum consists of a maze of membranous tubes and vesicles continuous with the inner chloroplast membrane that extends into the internal stromal fluid of the chloroplast. Its purpose is thought to be to increase the chloroplast's surface area for cross-membrane transport between its stroma and the cell cytoplasm. The small vesicles sometimes observed may serve as transport vesicles to shuttle stuff between the thylakoids and the outside of the chloroplast.

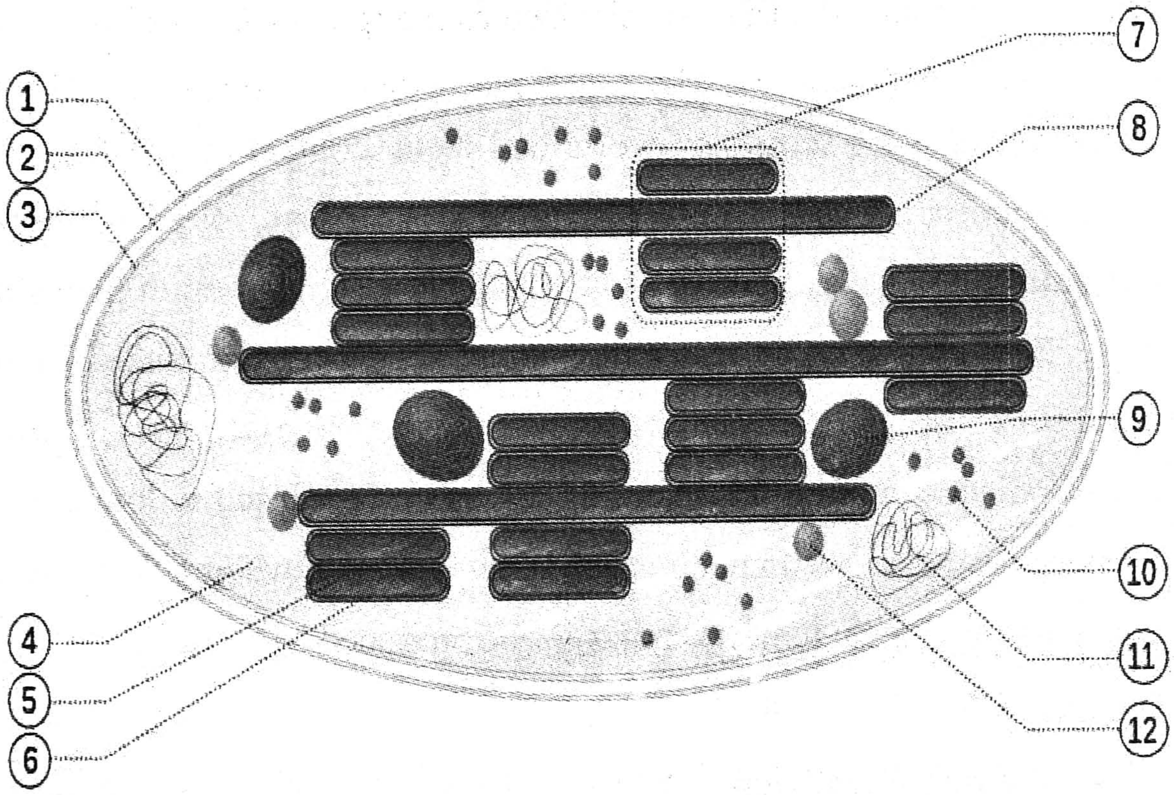
Chloroplast shapes

The chloroplasts of most higher plants are usually lens-shaped. Chloroplasts with quite different shapes occur in algae, such as a net (e.g., *Oedogonium*), a cup (e.g., *Chlamydomonas*), a ribbon-like spiral around the edges

of the cell (e.g., *Spirogyra*), or slightly twisted bands at the cell edges (e.g., *Sirogonium*). Some algae have two chloroplasts in each cell; they are star-shaped in *Zygnema*, or may follow the shape of half the cell in order Desmidiales.

Chloroplast colors

All chloroplasts contain the green pigment chlorophyll a, but not all chloroplasts are green because accessory pigments may be present that can change or override the green color.



Chloroplast ultrastructure:

1 outer membrane

2 intermembrane space

3 inner membrane (1+2+3: envelope)

4 stroma (aqueous fluid)

5 thylakoid lumen (inside of thylakoid)

6 thylakoid membrane

7 granum (stack of thylakoids)

8 thylakoid (lamella)

9 starch

10 ribosome

11 plastidial DNA

12 plastoglobule (drop of lipids)

Ultrastructure and functions of Mitochondrion

Mitochondria are the special protoplasmic organelles distributed in the cytoplasm of eukaryotic cells. Mitochondria contain the biochemical machinery involved in cellular respiration which take energy from breakdown of glucose and produce energy-rich ATP molecule which fuel the biochemecal reactions in the rest of the cell. Hence, mitochondria are described as the **'power houses'** of cells. Mitochondria were first observed by Altmann (1894). Hogeboom (1948) discovered that mitochondria are the site of aerobic respiration.

Ultrastructure structure

Mitochondria are microscopic and granular or cylindrical,

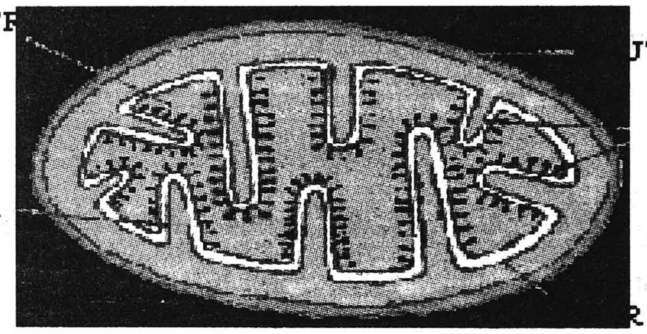
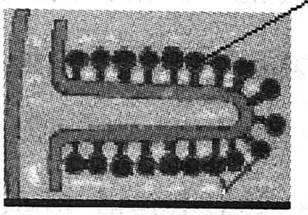


Figure A mitochondrion in section

Ultrastructure of a mitochondrion

And are bound by two lipo-protein membranes called outer and inner membranes. These are separated from each other by the **inter-membrane space.** The outer membrane is smooth, while the inner membrane is highly folded. The folds are called **cristae**, and project into the inner space of the mitochondrion and appear finger-shaped in the section.

The internal space of the mitochondrion (enclosed within the folded inner membrane) is filled with dense proteinaceous material and is called **the matrix**. It contains small **ribosomes** and **mitochondrial DNA** (DNA) invited in the replication of mitochondria. The matrix also contains all the enzymes required for the **Kreb's cycle reactions** during the second phase of aerobic respiration.



Structure of cristae membrane: The inner surface of the cristae membrane (i.e. the surface towards the matrix) is covered with numerous (infinite) stalked particles. These are called F_1 particles, elementary particles or sub units. These particles project into the matrix. Each F_1 particle has three parts, viz., the head piece, the stalk and the base piece. The respiratory chain is situated in the cristae membrane where the F_1 particles are present. The chain consists of enzymes and co-enzymes which form the electron transport system (ETS) in the mitochondrion. These enzymes and co-enzymes of the ETS act as the electron acceptors in the aerobic respiration reactions oxidative phosphorylation).

Outer Membrane

A mitochondrion contains outer and inner membranes composed of phospholipid bilayers and proteins.[6] The two membranes have different properties. Because of this double-membraned organization, there are five distinct parts to a mitochondrion. They are:

1. the outer mitochondrial membrane,
2. the intermembrane space (the space between the outer and inner membranes),
3. the inner mitochondrial membrane,
4. the cristae space (formed by infoldings of the inner membrane), and
5. the matrix (space within the inner membrane).

The outer mitochondrial membrane, which encloses the entire organelle, has a protein-to-phospholipid ratio similar to that of the eukaryotic plasma membrane (about 1:1 by weight). It contains large numbers of integral proteins called *porins*. These porins form channels that allow molecules 5000 Daltons or less in molecular weight to freely diffuse from one side of the membrane to the other. Larger proteins can enter the mitochondrion if a signaling sequence at their N-

terminus binds to a large multisubunit protein called translocase of the outer membrane, which then actively moves them across the membrane. Disruption of the outer membrane permits proteins in the intermembrane space to leak into the cytosol, leading to certain cell death.[16] The mitochondrial outer membrane can associate with the endoplasmic reticulum (ER) membrane, in a structure called MAM (mitochondria-associated ER-membrane). This is important in the ER-mitochondria calcium signaling and involved in the transfer of lipids between the ER and mitochondria.

Intermembrane space

The intermembrane space is the space between the outer membrane and the inner membrane. It is also known as Perimitochondrial space. Because the outer membrane is freely permeable to small molecules, the concentrations of small molecules such as ions and sugars in the intermembrane space is the same as the cytosol. However, large proteins must have a specific signaling sequence to be transported across the outer membrane, so the protein composition of this space is different from the protein composition of the cytosol. One protein that is localized to the intermembrane space in this way is cytochrome c.

Inner membrane

The inner mitochondrial membrane contains proteins with five types of functions:

1. Those that perform the redox reactions of oxidative phosphorylation
2. ATP synthase, which generates ATP in the matrix
3. Specific transport proteins that regulate metabolite passage into and out of the matrix
4. Protein import machinery.
5. Mitochondria fusion and fission protein.

It contains more than 151 different polypeptides, and has a very high protein-to-phospholipid ratio (more than 3:1 by weight, which is about 1 protein for 15 phospholipids). The inner membrane is home to around 1/5 of the total protein in a mitochondrion. In addition, the inner membrane is rich in an unusual phospholipid, cardiolipin. This phospholipid was originally discovered in cow

hearts in 1942, and is usually characteristic of mitochondrial and bacterial plasma membranes. Cardiolipin contains four fatty acids rather than two, and may help to make the inner membrane impermeable. Unlike the outer membrane, the inner membrane doesn't contain porins, and is highly impermeable to all molecules. Almost all ions and molecules require special membrane transporters to enter or exit the matrix. Proteins are ferried into the matrix via the translocase of the inner membrane (TIM) complex or via Oxa1. In addition, there is a membrane potential across the inner membrane, formed by the action of the enzymes of the electron transport chain.

Cristae

The inner mitochondrial membrane is compartmentalized into numerous cristae, which expand the surface area of the inner mitochondrial membrane, enhancing its ability to produce ATP. For typical liver mitochondria, the area of the inner membrane is about five times as great as the outer membrane. This ratio is variable and mitochondria from cells that have a greater demand for ATP, such as muscle cells, contain even more cristae. These folds are studded with small round bodies known as F_1 particles or oxysomes. These are not simple random folds but rather invaginations of the inner membrane, which can affect overall chemiosmotic function. One recent mathematical modeling study has suggested that the optical properties of the cristae in filamentous mitochondria may affect the generation and propagation of light within the tissue.

Matrix

The matrix is the space enclosed by the inner membrane. It contains about 2/3 of the total protein in a mitochondrion.[6] The matrix is important in the production of ATP with the aid of the ATP synthase contained in the inner membrane. The matrix contains a highly concentrated mixture of hundreds of enzymes, special mitochondrial ribosomes, tRNA, and several copies of the mitochondrial DNA genome. Of the enzymes, the major functions include oxidation of pyruvate and fatty acids, and the citric acid cycle. Mitochondria have their own genetic material, and the machinery to manufacture their own RNAs and proteins (*see: protein biosynthesis*). A published human mitochondrial DNA

sequence revealed 16,569 base pairs encoding 37 total genes: 22 tRNA, 2 rRNA, and 13 peptide genes.[21] The 13 mitochondrial peptides in humans are integrated into the inner mitochondrial membrane, along with proteins encoded by genes that reside in the host cell's nucleus.

Mitochondria-associated ER membrane (MAM)

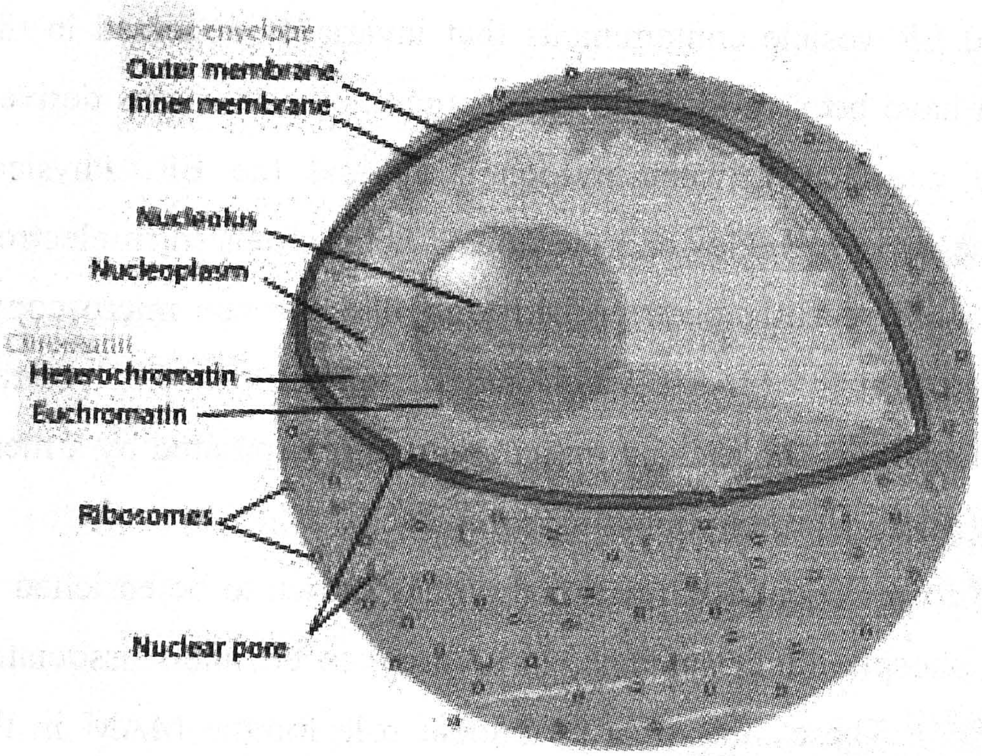
The mitochondria-associated ER membrane (MAM) is another structural element that is increasingly recognized for its critical role in cellular physiology and homeostasis. Once considered a technical snag in cell fractionation techniques, the alleged ER vesicle contaminants that invariably appeared in the mitochondrial fraction have been re-identified as membranous structures derived from the MAM—the interface between mitochondria and the ER. Physical coupling between these two organelles had previously been observed in electron micrographs and has more recently been probed with fluorescence microscopy. Such studies estimate that at the MAM, which may comprise up to 20% of the mitochondrial outer membrane, the ER and mitochondria are separated by a mere 10-25 nm and held together by protein tethering complexes.

Purified MAM from subcellular fractionation has shown to be enriched in enzymes involved in phospholipid exchange, in addition to channels associated with Ca^{2+} signaling.[22][24] These hints of a prominent role for the MAM in the regulation of cellular lipid stores and signal transduction have been borne out, with significant implications for mitochondrial-associated cellular phenomena, as discussed below. Not only has the MAM provided insight into the mechanistic basis underlying such physiological processes as intrinsic apoptosis and the propagation of calcium signalling, but it also favours a more refined view of the mitochondria. Though often seen as static, isolated 'powerhouses' hijacked for cellular metabolism through an ancient endosymbiotic event, the evolution of the MAM underscores the extent to which mitochondria have been integrated into overall cellular physiology, with intimate physical and functional coupling to the endomembrane system.

UNIT –2 STRUCTURES AND FUNCTION OF ORGANELLES

Structure

The nucleus is the largest cellular organelle in animals. In mammalian cells, the average diameter of the nucleus is approximately 6 micrometers (μm), which occupies about 10% of the total cell volume. The viscous liquid within it is called nucleoplasm, and is similar in composition to the cytosol found outside the nucleus. It appears as a dense, roughly spherical organelle.



The eukaryotic cell nucleus. Visible in this diagram are the ribosome-studded double membranes of the nuclear envelope, the DNA (complexed as chromatin), and the nucleolus. Within the cell nucleus is a viscous liquid called nucleoplasm, similar to the cytoplasm found outside the nucleus.

The nuclear envelope and pores

The nuclear envelope, otherwise known as nuclear membrane, consists of two cellular membranes, an inner and an outer membrane, arranged parallel to one another and separated by 10 to 50 nanometers (nm). The nuclear envelope completely encloses the nucleus and separates the cell's genetic material from the surrounding cytoplasm, serving as a barrier to prevent macromolecules from diffusing freely between the nucleoplasm and the cytoplasm. The outer nuclear

membrane is continuous with the membrane of the rough endoplasmic reticulum (RER), and is similarly studded with ribosomes. The space between the membranes is called the perinuclear space and is continuous with the RER lumen. Nuclear pores, which provide aqueous channels through the envelope, are composed of multiple proteins, collectively referred to as nucleoporins. The pores are about 125 million daltons in molecular weight and consist of around 50 (in yeast) to several hundred proteins (in vertebrates). The pores are 100 nm in total diameter; however, the gap through which molecules freely diffuse is only about 9 nm wide, due to the presence of regulatory systems within the center of the pore. This size allows the not-free passage of small water-soluble molecules while preventing larger molecules, such as nucleic acids and larger proteins, from inappropriately entering or exiting the nucleus. These large molecules must be actively transported into the nucleus instead. The nucleus of a typical mammalian cell will have about 3000 to 4000 pores throughout its envelope, each of which contains a donut-shaped, eightfold-symmetric ring-shaped structure at a position where the inner and outer membranes fuse. Attached to the ring is a structure called the *nuclear basket* that extends into the nucleoplasm, and a series of filamentous extensions that reach into the cytoplasm. Both structures serve to mediate binding to nuclear transport proteins.

Most proteins, ribosomal subunits, and some DNAs are transported through the pore complexes in a process mediated by a family of transport factors known as karyopherins. Those karyopherins that mediate movement into the nucleus are also called importins, whereas those that mediate movement out of the nucleus are called exportins. Most karyopherins interact directly with their cargo, although some use adaptor proteins. Steroid hormones such as cortisol and aldosterone, as well as other small lipid-soluble molecules involved in intercellular signaling, can diffuse through the cell membrane and into the cytoplasm, where they bind nuclear receptor proteins that are trafficked into the nucleus. There they serve as transcription factors when bound to their ligand; in the absence of ligand, many such receptors function as histone deacetylases that repress gene expression.

Nuclear lamina

In cells, two networks of intermediate filaments provide the nucleus with mechanical support: The nuclear lamina forms an organized meshwork on the internal face of the envelope, while less organized support is provided on the cytosolic face of the envelope. Both systems provide structural support for the nuclear envelope and anchoring sites for chromosomes and nuclear pores.

The nuclear lamina is composed mostly of lamin proteins. Like all proteins, lamins are synthesized in the cytoplasm and later transported into the nucleus interior, where they are assembled before being incorporated into the existing network of nuclear lamina. Lamins found on the cytosolic face of the membrane, such as emerin and nesprin, bind to the cytoskeleton to provide structural support. Lamins are also found inside the nucleoplasm where they form another regular structure, known as the *nucleoplasmic veil*, that is visible using fluorescence microscopy. The actual function of the veil is not clear, although it is excluded from the nucleolus and is present during interphase. Lamin structures that make up the veil, such as LEM3, bind chromatin and disrupting their structure inhibits transcription of protein-coding genes.

Like the components of other intermediate filaments, the lamin monomer contains an alpha-helical domain used by two monomers to coil around each other, forming a dimer structure called a coiled coil. Two of these dimer structures then join side by side, in an antiparallel arrangement, to form a tetramer called a *protofilament*. Eight of these protofilaments form a lateral arrangement that is twisted to form a ropelike *filament*. These filaments can be assembled or disassembled in a dynamic manner, meaning that changes in the length of the filament depend on the competing rates of filament addition and removal.

Mutations in lamin genes leading to defects in filament assembly are known as *laminopathies*. The most notable laminopathy is the family of diseases known as progeria, which causes the appearance of premature aging in its sufferers. The exact mechanism by which the associated biochemical changes give rise to the aged phenotype is not well understood.

Chromosomes

The cell nucleus contains the majority of the cell's genetic material in the form of multiple linear DNA molecules organized into structures called chromosomes. Each human cell contains 2m of DNA. During most of the cell cycle these are organized in a DNA-protein complex known as chromatin, and during cell division the chromatin can be seen to form the well-defined chromosomes familiar from a karyotype. A small fraction of the cell's genes are located instead in the mitochondria.

There are two types of chromatin. Euchromatin is the less compact DNA form, and contains genes that are frequently expressed by the cell. The other type, heterochromatin, is the more compact form, and contains DNA that are infrequently transcribed. This structure is further categorized into *facultative* heterochromatin, consisting of genes that are organized as heterochromatin only in certain cell types or at certain stages of development, and *constitutive* heterochromatin that consists of chromosome structural components such as telomeres and centromeres. During interphase the chromatin organizes itself into discrete individual patches, called *chromosome territories.* Active genes, which are generally found in the euchromatic region of the chromosome, tend to be located towards the chromosome's territory boundary.

Antibodies to certain types of chromatin organization, in particular, nucleosomes, have been associated with a number of autoimmune diseases, such as systemic lupus erythematosus. These are known as anti-nuclear antibodies (ANA) and have also been observed in concert with multiple sclerosis as part of general immune system dysfunction. As in the case of progeria, the role played by the antibodies in inducing the symptoms of autoimmune diseases is not obvious.
The main function of the cell nucleus is to control gene expression and mediate the replication of DNA during the cell cycle. The nucleus provides a site for genetic transcription that is segregated from the location of translation in the cytoplasm, allowing levels of gene regulation that are not available to prokaryotes.

Cell compartmentalization

The nuclear envelope allows the nucleus to control its contents, and separate them from the rest of the cytoplasm where necessary. This is important for controlling processes on either side of the nuclear membrane. In most cases where a cytoplasmic process needs to be restricted, a key participant is removed to the nucleus, where it interacts with transcription factors to downregulate the production of certain enzymes in the pathway. This regulatory mechanism occurs in the case of glycolysis, a cellular pathway for breaking down glucose to produce energy. Hexokinase is an enzyme responsible for the first the step of glycolysis, forming glucose-6-phosphate from glucose. At high concentrations of fructose-6-phosphate, a molecule made later from glucose-6-phosphate, a regulator protein removes hexokinase to the nucleus, where it forms a transcriptional repressor complex with nuclear proteins to reduce the expression of genes involved in glycolysis.

In order to control which genes are being transcribed, the cell separates some transcription factor proteins responsible for regulating gene expression from physical access to the DNA until they are activated by other signaling pathways. This prevents even low levels of inappropriate gene expression. For example, in the case of NF-κB-controlled genes, which are involved in most inflammatory responses, transcription is induced in response to a signal pathway such as that initiated by the signaling molecule TNF-α, binds to a cell membrane receptor, resulting in the recruitment of signalling proteins, and eventually activating the transcription factor NF-κB. A nuclear localisation signal on the NF-κB protein allows it to be transported through the nuclear pore and into the nucleus, where it stimulates the transcription of the target genes.

The compartmentalization allows the cell to prevent translation of unspliced mRNA. Eukaryotic mRNA contains introns that must be removed before being translated to produce functional proteins. The splicing is done inside the nucleus before the mRNA can be accessed by ribosomes for translation. Without the nucleus, ribosomes would translate newly transcribed (unprocessed) mRNA, resulting in misformed and nonfunctional proteins.

Since the nucleus is the site of transcription, it also contains a variety of proteins that either directly mediate transcription or are involved in regulating the process. These proteins include helicases, which unwind the double-stranded DNA molecule to facilitate access to it, RNA polymerases, which synthesize the growing RNA molecule, topoisomerases, which change the amount of supercoiling in DNA, helping it wind and unwind, as well as a large variety of transcription factors that regulate expression.

Processing of pre-mRNA

Newly synthesized mRNA molecules are known as primary transcripts or pre-mRNA. They must undergo post-transcriptional modification in the nucleus before being exported to the cytoplasm; mRNA that appears in the cytoplasm without these modifications is degraded rather than used for protein translation. The three main modifications are 5' capping, 3' polyadenylation, and RNA splicing. While in the nucleus, pre-mRNA is associated with a variety of proteins in complexes known as heterogeneous ribonucleoprotein particles (hnRNPs). Addition of the 5' cap occurs co-transcriptionally and is the first step in post-transcriptional modification. The 3' poly-adenine tail is only added after transcription is complete.

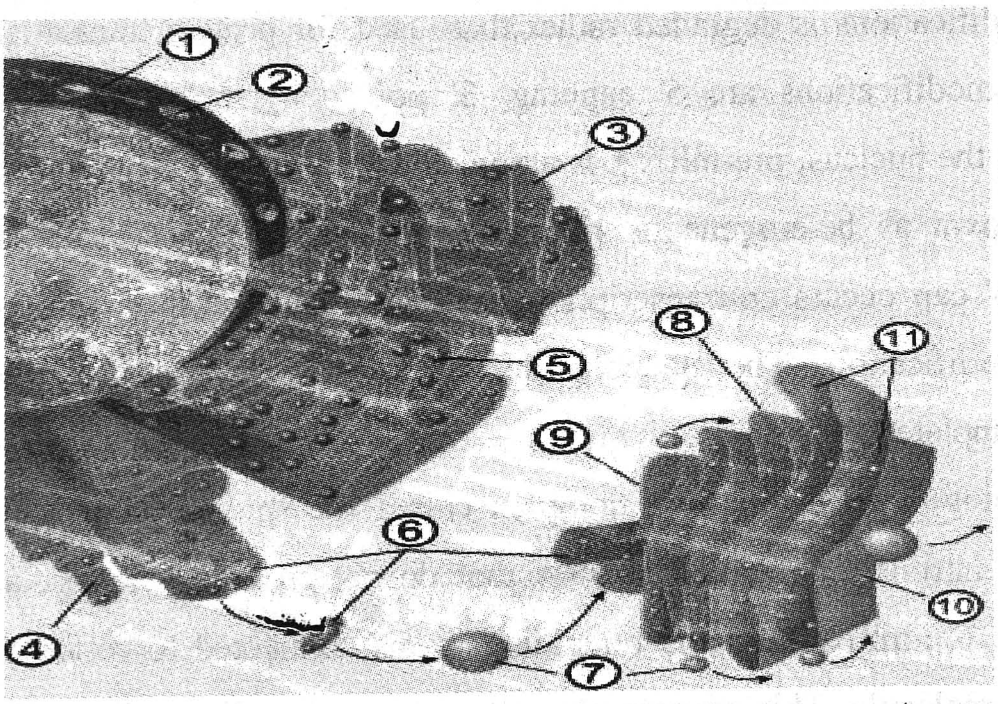
RNA splicing, carried out by a complex called the spliceosome, is the process by which introns, or regions of DNA that do not code for protein, are removed from the pre-mRNA and the remaining exons connected to re-form a single continuous molecule. This process normally occurs after 5' capping and 3' polyadenylation but can begin before synthesis is complete in transcripts with many exons.[5] Many pre-mRNAs, including those encoding antibodies, can be spliced in multiple ways to produce different mature mRNAs that encode different protein sequences. This process is known as alternative splicing, and allows production of a large variety of proteins from a limited amount of DNA.

Endoplasmic Reticulum (ER)

The endoplasmic reticulum (ER) is an organelle of cells in eukaryotic organisms that forms an interconnected network of membrane vesicles. According to the structure the endoplasmic reticulum is classified into two types, that is,

rough endoplasmic reticulum (RER) and smooth endoplasmic reticulum (SER). The rough endoplasmic reticulum is studded with *ribosomes* on the *cytosolic* face. These are the sites of *protein synthesis.* The rough endoplasmic reticulum is predominantly found in *hepatocytes* where protein synthesis occurs actively. The smooth endoplasmic reticulum is a smooth network without the ribosomes. The smooth endoplasmic reticulum is concerned with *lipid metabolism, carbohydrate metabolism* and *detoxification.* The smooth endoplasmic reticulum is abundantly found in mammalian liver and gonad cells. The lacey membranes of the endoplasmic reticulum were first seen by Keith R. Porter, Albert Claude, and Ernest F. Fullam in the year 1945.

Structure



1 Nucleus **2** Nuclear pore **3** Rough endoplasmic reticulum (RER) **4** Smooth endoplasmic reticulum (SER) **5** Ribosome on the rough ER **6** Proteins that are transported **7** Transport vesicle **8** Golgi apparatus **9** Cis face of the Golgi apparatus **10** Trans face of the Golgi apparatus **11** Cisternae of the Golgi apparatus. The general structure of the endoplasmic reticulum is a membranous network of cisternae (sac-like structures) held together by the cytoskeleton. The phospholipid membrane encloses a space, the cisternal space (or lumen), which is continuous with the perinuclear space but separate from the cytosol. The functions of the endoplasmic reticulum vary greatly depending on its cell type, cell function, and cell needs. The ER can even modify to change over time in response to cell

needs. The three most common varieties are called *rough endoplasmic reticulum*, *smooth endoplasmic reticulum*, and *sarcoplasmic reticulum*.

The quantity of RER and SER in a cell can slowly interchange from one type to the other, depending on changing metabolic needs. Transformation can include embedment of new proteins in membrane as well as structural changes. Massive changes may also occur in protein content without noticeable structural changes.

Rough endoplasmic reticulum

The surface of the rough endoplasmic reticulum (often abbreviated RER) is studded with protein-manufacturing ribosomes giving it a "rough" appearance (hence its name). The binding site of the ribosome on the rough endoplasmic reticulum is the translocon. However, the ribosomes bound to it at any one time are not a stable part of this organelle's structure as they are constantly being bound and released from the membrane. A ribosome binds to the endoplasmic reticulum only when it begins to synthesize a protein destined for the secretory pathway. There, a ribosome in the cytosol begins synthesizing a protein until a signal recognition particle recognizes the signal peptide of 5–30 hydrophobic amino acids, sometimes preceded by a positively charged amino acid. This signal sequence allows the recognition particle to bind to the ribosome, causing the ribosome to bind to the rough endoplasmic reticulum and pass the new protein through the rough endoplasmic reticulum membrane. The signal peptide is then cleaved off within the lumen of the ER by a signal peptidase. Ribosomes at this point may be released back into the cytosol, however non-translating ribosomes are also known to stay associated with translocons.

The membrane of the rough endoplasmic reticulum forms large double membrane sheets that are located near, and continuous with the outer layer of the nuclear envelope. Although there is no continuous membrane between the endoplasmic reticulum and the Golgi apparatus, membrane-bound vesicles shuttle proteins between these two compartments. Vesicles are surrounded by coating proteins called COPI and COPII. COPII targets vesicles to the golgi apparatus and COPI marks them to be brought back to the rough endoplasmic reticulum. The

rough endoplasmic reticulum works in concert with the golgi complex to target new proteins to their proper destinations. A second method of transport out of the endoplasmic reticulum involves areas called membrane contact sites, where the membranes of the endoplasmic reticulum and other organelles are held closely together, allowing the transfer of lipids and other small molecules.

The rough endoplasmic reticulum is key in multiple functions:

- Manufacture of lysosomal enzymes with a mannose-6-phosphate marker added in the *cis*-Golgi network[*citation needed*]
- Manufacture of secreted proteins, either secreted constitutively with no tag or secreted in a regulatory manner involving clathrin and paired basic amino acids in the signal peptide.
- Integral membrane proteins that stay embedded in the membrane as vesicles exit and bind to new membranes. Rab proteins are key in targeting the membrane; SNAP and SNARE proteins are key in the fusion event.
- Initial glycosylation as assembly continues. This is N-linked (O-linking occurs in the golgi).
    - N-linked glycosylation: If the protein is properly folded, glycosyltransferase recognizes the AA sequence NXS or NXT (with the S/T residue phosphorylated) and adds a 14-sugar backbone (2-*N*-acetylglucosamine, 9-branching mannose, and 3-glucose at the end) to the side-chain nitrogen of Asn.

Endoplasmic reticulum stress

Disturbances in redox regulation, calcium regulation, glucose deprivation, and viral infection can lead to endoplasmic reticulum stress. It is a state in which the folding of proteins slows leading to an increase in unfolded proteins. This stress is emerging as a potential cause of damage in hypoxia/ischemia, insulin resistance and other disorders.

Smooth endoplasmic reticulum

The smooth endoplasmic reticulum (SER) has functions in several metabolic processes. It synthesizes lipids, phospholipids and steroids—cells which secrete these products, such as those in the testes, ovaries, and skin oil glands have a great deal of smooth endoplasmic reticulum. It also carries out the metabolism of carbohydrates, drug detoxification, attachment of receptors on cell membrane proteins, and steroid metabolism. In muscle cells, it regulates calcium ion concentration. It is connected to the nuclear envelope. Smooth endoplasmic reticulum is found in a variety of cell types (both animal and plant), and it serves different functions in each. The smooth endoplasmic reticulum also contains the enzyme glucose-6-phosphatase, which converts glucose-6-phosphate to glucose, a step in gluconeogenesis. It consists of tubules that are located near the cell periphery. These tubes sometimes branch forming a network that is reticular in appearance. In some cells, there are dilated areas like the sacs of rough endoplasmic reticulum. The network of smooth endoplasmic reticulum allows increased surface area for the action or storage of key enzymes and the products of these enzymes.

Sarcoplasmic reticulum

The sarcoplasmic reticulum (SR), from the Greek *sarx*, ("flesh"), is smooth ER found in smooth and striated muscle. The only structural difference between this organelle and the smooth endoplasmic reticulum is the medley of proteins they have, both bound to their membranes and drifting within the confines of their lumens. This fundamental difference is indicative of their functions: The endoplasmic reticulum synthesizes molecules, while the sarcoplasmic reticulum stores and pumps calcium ions. The sarcoplasmic reticulum contains large stores of calcium, which it sequesters and then releases when the muscle cell is stimulated. It plays a major role in excitation-contraction coupling.

Functions

The endoplasmic reticulum serves many general functions, including the facilitation of protein folding and the transport of synthesized proteins in sacs called cisternae. Correct folding of newly made proteins is made possible by several endoplasmic reticulum chaperone proteins, including protein disulfide isomerase (PDI), ERp29, the Hsp70 family member Grp78, calnexin, calreticulin, and the peptidylpropyl isomerase family. Only properly folded proteins are transported from the rough ER to the Golgi complex.

Transport of proteins

Secretory proteins, mostly glycoproteins, are moved across the endoplasmic reticulum membrane. Proteins that are transported by the endoplasmic reticulum throughout the cell are marked with an address tag called a signal sequence. The N-terminus (one end) of a polypeptide chain (i.e., a protein) contains a few amino acids that work as an address tag, which are removed when the polypeptide reaches its destination. Proteins that are destined for places outside the endoplasmic reticulum are packed into transport vesicles and moved along the cytoskeleton toward their destination.

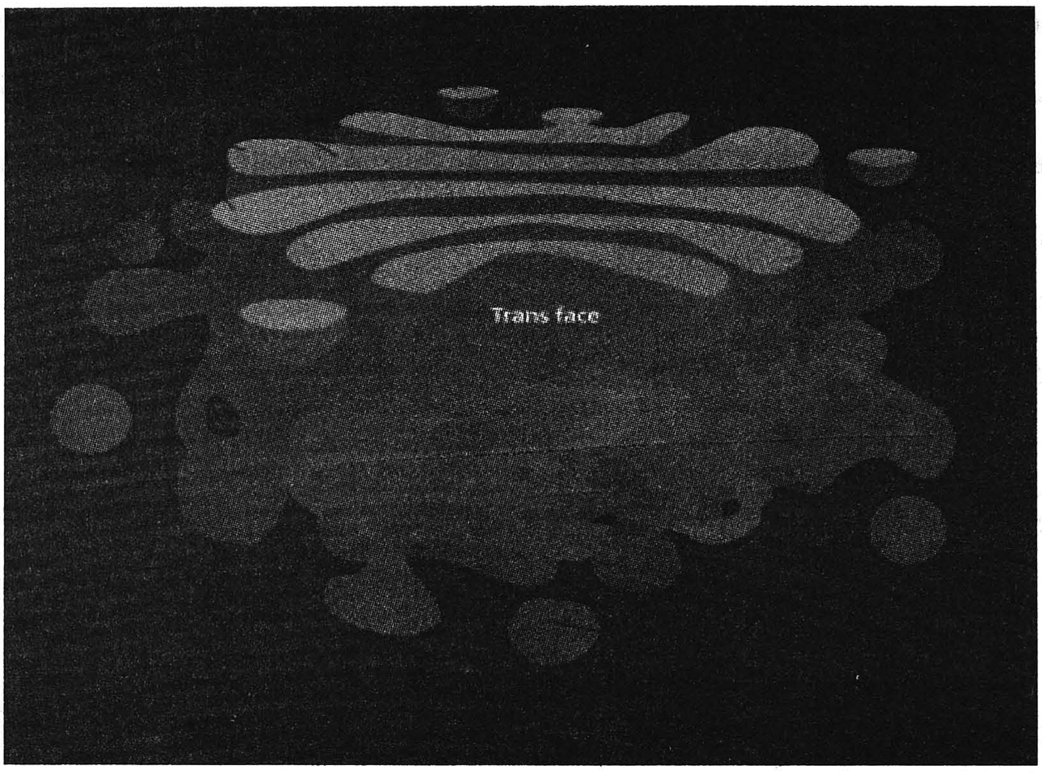
The endoplasmic reticulum is also part of a protein sorting pathway. It is, in essence, the transportation system of the eukaryotic cell. The majority of its resident proteins are retained within it through a retention motif. This motif is composed of four amino acids at the end of the protein sequence. The most common retention sequence is KDEL (*lys-asp-glu-leu*). However, variation on KDEL does occur and other sequences can also give rise to endoplasmic reticulum retention. It is not known whether such variation can lead to sub-ER localizations. There are three KDEL receptors in mammalian cells, and they have a very high degree of sequence identity. The functional differences between these receptors remain to be established.[*citation needed*]

Golgi apparatus,

The Golgi apparatus, also known as the Golgi complex, Golgi body, or simply the Golgi, is an organelle found in most eukaryotic cells. It was identified in 1897 by the Italian physician Camillo Golgi and named after him in 1898.

Part of the cellular endomembrane system, the Golgi apparatus packages proteins inside the cell before they are sent to their destination; it is particularly important in the processing of proteins for secretion.

Structure



Found within the cytoplasm of both plant and animal cells, the Golgi is composed of stacks of membrane-bound structures known as cisternae (singular: *cisterna*). An individual stack is sometimes called a dictyosome (from Greek *dictyon*: net + *soma*: body), especially in plant cells. A mammalian cell typically contains 40 to 100 stacks. Between four and eight cisternae are usually present in a stack; however, in some protists as many as sixty have been observed. Each cisterna comprises a flat, membrane enclosed disc that includes special Golgi enzymes which modify or help to modify cargo proteins that travel through it.

The cisternae stack has four functional regions: the cis-Golgi network, medial-Golgi, endo-Golgi, and trans-Golgi network. Vesicles from the endoplasmic reticulum (via the vesicular-tubular clusters) fuse with the network and subsequently progress through the stack to the trans Golgi network, where they are packaged and sent to their destination. Each region contains different enzymes which selectively modify the contents depending on where they reside. The cisternae also carry structural proteins important for their maintenance as flattened membranes which stack upon each other.

Function of a golgi body

Cells synthesize a large number of different macromolecules. The Golgi apparatus is integral in modifying, sorting, and packaging these macromolecules for cell secretion (exocytosis) or use within the cell. It primarily modifies proteins delivered from the rough endoplasmic reticulum but is also involved in the transport of lipids around the cell, and the creation of lysosomes. In this respect it can be thought of as similar to a post office; it packages and labels items which it then sends to different parts of the cell.

Enzymes within the cisternae are able to modify the proteins by addition of carbohydrates (glycosylation) and phosphates (phosphorylation). In order to do so, the Golgi imports substances such as nucleotide sugars from the cytosol. These modifications may also form a signal sequence which determines the final destination of the protein. For example, the Golgi apparatus adds a mannose-6-phosphate label to proteins destined for lysosomes.

The Golgi plays an important role in the synthesis of proteoglycans, which are molecules present in the extracellular matrix of animals. It is also a major site of carbohydrate synthesis. This includes the production of glycosaminoglycans (GAGs), long unbranched polysaccharides which the Golgi then attaches to a protein synthesised in the endoplasmic reticulum to form proteoglycans. Enzymes in the Golgi polymerize several of these GAGs via a xylose link onto the core protein. Another task of the Golgi involves the sulfation of certain molecules passing through its lumen via sulfotranferases that gain their sulfur molecule from a donor called PAPS. This process occurs on the GAGs of proteoglycans as well as on the core protein. Sulfation is generally performed in the trans-Golgi network. The level of sulfation is very important to the proteoglycans' signalling abilities as well as giving the proteoglycan its overall negative charge.

The phosphorylation of molecules requires that ATP is imported into the lumen of the Golgi[15] and utilised by resident kinases such as casein kinase 1 and casein kinase 2. One molecule that is phosphorylated in the Golgi is Apolipoprotein, which forms a molecule known as VLDL that is a constituent of

blood serum. It is thought that the phosphorylation of these molecules is important to help aid in their sorting for secretion into the blood serum.

The Golgi has a putative role in apoptosis, with several Bcl-2 family members localised there, as well as to the mitochondria. A newly characterized protein, GAAP (Golgi anti-apoptotic protein), almost exclusively resides in the Golgi and protects cells from apoptosis by an as-yet undefined mechanism.

Vesicular transport

The vesicles that leave the rough endoplasmic reticulum are transported to the *cis* face of the Golgi apparatus, where they fuse with the Golgi membrane and empty their contents into the lumen. Once inside the lumen, the molecules are modified, then sorted for transport to their next destinations. The Golgi apparatus tends to be larger and more numerous in cells that synthesize and secrete large amounts of substances; for example, the plasma B cells and the antibody-secreting cells of the immune system have prominent Golgi complexes.

Those proteins destined for areas of the cell other than either the endoplasmic reticulum or Golgi apparatus are moved towards the *trans* face, to a complex network of membranes and associated vesicles known as the *trans-Golgi network* (TGN). This area of the Golgi is the point at which proteins are sorted and shipped to their intended destinations by their placement into one of at least three different types of vesicles, depending upon the molecular marker they carry.

Ribosome

The **ribosome** (from *ribo*nucleic acid and the Greek *soma*, meaning "body") is a large and complex molecular machine, found within all living cells, that serves as the primary site of biological protein synthesis (translation). Ribosomes link amino acids together in the order specified by messenger RNA (mRNA) molecules. Ribosomes consist of two major subunits—the small ribosomal subunit reads the mRNA, while the large subunit joins amino acids to form a polypeptide chain. Each subunit is composed of one or more ribosomal RNA (rRNA) molecules and a variety of proteins.

The sequence of DNA encoding for a protein may be copied many times into messenger RNA (mRNA) chains of a similar sequence. Ribosomes can bind

to an mRNA chain and use it as a template for determining the correct sequence of amino acids in a particular protein. Amino acids are selected, collected and carried to the ribosome by transfer RNA (tRNA molecules), which enter one part of the ribosome and bind to the messenger RNA chain. The attached amino acids are then linked together by another part of the ribosome. Once the protein is produced, it can then 'fold' to produce a specific functional three-dimensional structure.

A ribosome is made from complexes of RNAs and proteins and is therefore a ribonucleoprotein. Each ribosome is divided into two subunits: the smaller subunit binds to the mRNA pattern, while the larger subunit binds to the tRNA and the amino acids. When a ribosome finishes reading an mRNA molecule, these two subunits split apart. Ribosomes are ribozymes, because the catalytic peptidyl transferase activity that links amino acids together is performed by the ribosomal RNA.

Ribosomes from bacteria, archaea and eukaryotes (the three domains of life on Earth) differ in their size, sequence, structure, and the ratio of protein to RNA. The differences in structure allow some antibiotics to kill bacteria by inhibiting their ribosomes, while leaving human ribosomes unaffected. In bacteria and archaea, more than one ribosome may move along a single mRNA chain at one time, each "reading" its sequence and producing a corresponding protein molecule. The ribosomes in the mitochondria of eukaryotic cells functionally resemble many features of those in bacteria, reflecting the likely evolutionary origin of mitochondria.[1][2]

Albert Claude, Christian de Duve, and George Emil Palade were jointly awarded the Nobel Prize in Physiology or Medicine, in 1974, for the discovery of the ribosomes.[3] The Nobel Prize in Chemistry 2009 was awarded to Venkatraman Ramakrishnan, Thomas A. Steitz and Ada E. Yonath for determining the detailed structure and mechanism of the ribosome.[4]

Description

Ribosomes consist of two subunits that fit together (Figure 2) and work as one to translate the mRNA into a polypeptide chain during protein synthesis.

Because they are formed from two subunits of non-equal size, they are slightly longer in the axis than in diameter. Prokaryotic ribosomes are around 20 nm (200 Å) in diameter and are composed of 65% ribosomal RNA and 35% ribosomal proteins. Eukaryotic ribosomes are between 25 and 30 nm (250–300 Å) in diameter and the ratio of rRNA to protein is close to 1. Bacterial subunits consist of one or two and eukaryotic of one or three very large RNA molecules (known as ribosomal RNA or rRNA) and multiple smaller protein molecules. Crystallographic work has shown that there are no ribosomal proteins close to the reaction site for polypeptide synthesis. This proves that the protein components of ribosomes do not directly participate in peptide bond formation catalysis, but rather suggests that these proteins act as a scaffold that may enhance the ability of rRNA to synthesize protein.

Ribosomes translate polypeptide chains (e.g., proteins) from the genetic instructions held within messenger RNA, using amino acids delivered by transfer RNA (tRNA). Free ribosomes are suspended in the cytosol (the semi-fluid portion of the cytoplasm); others are bound to the rough endoplasmic reticulum, giving it the appearance of roughness and thus its name, or to the nuclear envelope. Although catalysis of the peptide bond involves the C2 hydroxyl of RNA's P-site (see Function section below) adenosine in a protein shuttle mechanism, other steps in protein synthesis (such as translocation) are caused by changes in protein conformations.Since their catalytic core is made of RNA, ribosomes are classified as "ribozymes," and it is thought that they might be remnants of the RNA world.

Ribosomes are sometimes referred to as organelles, but the use of the term *organelle* is often restricted to describing sub-cellular components that include a phospholipid membrane, which ribosomes, being entirely particulate, do not. For this reason, ribosomes may sometimes be described as "non-membranous organelles".

Ribosomes were first observed in the mid-1950s by Romanian cell biologist George Emil Palade using an electron microscope as dense particles or granules for which, in 1974, he would win a Nobel Prize. The term "ribosome" was proposed by scientist Richard B. Roberts in 1958:

During the course of the symposium a semantic difficulty became apparent. To some of the participants, "microsomes" mean the ribonucleoprotein particles of the microsome fraction contaminated by other protein and lipid material; to others, the microsomes consist of protein and lipid contaminated by particles. The phrase “microsomal particles” does not seem adequate, and “ribonucleoprotein particles of the microsome fraction” is much too awkward. During the meeting, the word "ribosome" was suggested, which has a very satisfactory name and a pleasant sound. The present confusion would be eliminated if “ribosome” were adopted to designate ribonucleoprotein particles in sizes ranging from 35 to 100S.

The structure and function of the ribosomes and associated molecules, known as the *translational apparatus*, has been of research interest since the mid-twentieth century and is a very active field of study today.

Biogenesis

In bacterial cells, ribosomes are synthesized in the cytoplasm through the transcription of multiple ribosome gene operons. In eukaryotes, the process takes place both in the cell cytoplasm and in the nucleolus, which is a region within the cell nucleus. The assembly process involves the coordinated function of over 200 proteins in the synthesis and processing of the four rRNAs, as well as assembly of those rRNAs with the ribosomal proteins.

Ribosome locations

Ribosomes are classified as being either "free" or "membrane-bound". A ribosome translating a protein that is secreted into the endoplasmic reticulum. Free and membrane-bound ribosomes differ only in their spatial distribution; they are identical in structure. Whether the ribosome exists in a free or membrane-bound state depends on the presence of an ER-targeting signal sequence on the protein being synthesized, so an individual ribosome might be membrane-bound when it is making one protein, but free in the cytosol when it makes another protein.

Free ribosomes

Free ribosomes can move about anywhere in the cytosol, but are excluded from the cell nucleus and other organelles. Proteins that are formed from free

ribosomes are released into the cytosol and used within the cell. Since the cytosol contains high concentrations of glutathione and is, therefore, a reducing environment, proteins containing disulfide bonds, which are formed from oxidized cysteine residues, cannot be produced in this compartment.

Membrane-bound ribosomes

When a ribosome begins to synthesize proteins that are needed in some organelles, the ribosome making this protein can become "membrane-bound". In eukaryotic cells this happens in a region of the endoplasmic reticulum (ER) called the "rough ER". The newly produced polypeptide chains are inserted directly into the ER by the ribosome undertaking vectorial synthesis and are then transported to their destinations, through the secretary pathway. Bound ribosomes usually produce proteins that are used within the plasma membrane or are expelled from the cell via *exocytosis*.

Atomic structure of the 30S Subunit from *Thermus thermophilus*. Proteins are shown in blue and the single RNA chain in orange. The ribosomal subunits of prokaryotes and eukaryotes are quite similar. The unit of measurement is the Svedberg unit, a measure of the rate of sedimentation in centrifugation rather than size, and this accounts for why fragment names do not add up (70S is made of 50S and 30S).

Prokaryotes have 70S ribosomes, each consisting of a small (30S) and a large (50S) subunit. Their small subunit has a 16S RNA subunit (consisting of 1540 nucleotides) bound to 21 proteins. The large subunit is composed of a 5S RNA subunit (120 nucleotides), a 23S RNA subunit (2900 nucleotides) and 31 proteins. Affinity label for the tRNA binding sites on the E. coli ribosome allowed the identification of A and P site proteins most likely associated with the peptidyltransferase activity; labelled proteins are L27, L14, L15, L16, L2; at least L27 is located at the donor site, as shown by E. Collatz and A.P. Czernilofsky. Additional research has demonstrated that the S1 and S21 proteins, in association with the 3'-end of 16S ribosomal RNA, are involved in the initiation of translation.

Eukaryotes have 80S ribosomes, each consisting of a small (40S) and large (60S) subunit. Their 40S subunit has an 18S RNA (1900 nucleotides) and 33

proteins. The large subunit is composed of a 5S RNA (120 nucleotides), 28S RNA (4700 nucleotides), a 5.8S RNA (160 nucleotides) subunits and 46 proteins.[11][15][17] During 1977, Czernilofsky published research that used affinity labeling to identify tRNA-binding sites on rat liver ribosomes. Several proteins, including L32/33, L36, L21, L23, L28/29 and L13 were implicated as being at or near the peptidyl transferase center.

The ribosomes found in chloroplasts and mitochondria of eukaryotes also consist of large and small subunits bound together with proteins into one 70S particle. These organelles are believed to be descendants of bacteria (see Endosymbiotic theory) and as such their ribosomes are similar to those of bacteria.

The various ribosomes share a core structure, which is quite similar despite the large differences in size. Much of the RNA is highly organized into various tertiary structural motifs, for example pseudoknots that exhibit coaxial stacking. The extra RNA in the larger ribosomes is in several long continuous insertions, such that they form loops out of the core structure without disrupting or changing it. All of the catalytic activity of the ribosome is carried out by the RNA; the proteins reside on the surface and seem to stabilize the structure.

The differences between the bacterial and eukaryotic ribosomes are exploited by pharmaceutical chemists to create antibiotics that can destroy a bacterial infection without harming the cells of the infected person. Due to the differences in their structures, the bacterial 70S ribosomes are vulnerable to these antibiotics while the eukaryotic 80S ribosomes are not. Even though mitochondria possess ribosomes similar to the bacterial ones, mitochondria are not affected by these antibiotics because they are surrounded by a double membrane that does not easily admit these antibiotics into the organelle.

Atomic structure of the 50S Subunit from *Haloarcula marismortui*. Proteins are shown in blue and the two RNA chains in orange and yellow.[22] The small patch of green in the center of the subunit is the active site.

The general molecular structure of the ribosome has been known since the early 1970s. In the early 2000s the structure has been achieved at high resolutions, on the order of a few Å.

The first papers giving the structure of the ribosome at atomic resolution were published almost simultaneously in late 2000. The 50S (large prokaryotic) subunit was determined from the archaeons *Haloarcula marismortui* and *Deinococcus radiodurans*, and the structure of the 30S subunit was determined from *Thermus thermophilus*. These structural studies were awarded the Nobel Prize in Chemistry in 2009. Early the next year (May 2001) these coordinates were used to reconstruct the entire *T. thermophilus* 70S particle at 5.5 Å resolution.

Two papers were published in November 2005 with structures of the *Escherichia coli* 70S ribosome. The structures of a vacant ribosome were determined at 3.5-Å resolution using x-ray crystallography. Then, two weeks later, a structure based on cryo-electron microscopy was published, which depicts the ribosome at 11–15 Å resolution in the act of passing a newly synthesized protein strand into the protein-conducting channel.

The first atomic structures of the ribosome complexed with tRNA and mRNA molecules were solved by using X-ray crystallography by two groups independently, at 2.8 Å and at 3.7 Å. These structures allow one to see the details of interactions of the *Thermus thermophilus* ribosome with mRNA and with tRNAs bound at classical ribosomal sites. Interactions of the ribosome with long mRNAs containing Shine-Dalgarno sequences were visualized soon after that at 4.5- to 5.5-Å resolution.

More recently, the first complete atomic structure of the eukaryotic 80S ribosome from the yeast *Saccharomyces cerevisiae* was obtained by crystallography. The model reveals the architecture of eukaryote-specific elements and their interaction with the universally conserved core. At the same time, the complete model of a eukaryotic 40S ribosomal structure in *Tetrahymena thermophila* was published and described the structure of the 40S subunit as well as much about the 40S subunit's interaction with F1 during translation initiation.

Similarly, the eukaryotic 60S subunit structure was also determined from *Tetrahymena thermophila* in complex with F6.

Function

Ribosomes are the workhorses of protein biosynthesis, the process of translating mRNA into protein. The mRNA comprises a series of codons that dictate to the ribosome the sequence of the amino acids needed to make the protein. Using the mRNA as a template, the ribosome traverses each codon (3 nucleotides) of the mRNA, pairing it with the appropriate amino acid provided by an aminoacyl-tRNA. aminoacyl-tRNA contains a complementary anticodon on one end and the appropriate amino acid on the other. The small ribosomal subunit, typically bound to an aminoacyl-tRNA containing the amino acid methionine, binds to an AUG codon on the mRNA and recruits the large ribosomal subunit. The ribosome contains three RNA binding sites, designated A, P and E. The A site binds an aminoacyl-tRNA; the P site binds a peptidyl-tRNA (a tRNA bound to the peptide being synthesized); and the E site binds a free tRNA before it exits the ribosome. Protein synthesis begins at a start codon AUG near the 5' end of the mRNA. mRNA binds to the P site of the ribosome first. The ribosome is able to identify the start codon by use of the Shine-Dalgarno sequence of the mRNA in prokaryotes and Kozak box in eukaryotes.

IDENTIFICATION OF DNA AS GENETIC MATERIAL

Deoxyribonucleic acid (DNA) has been proved to be the genetic material by two historical experiments conducted independently by Griffith and Avery Mc Leod and Mc Carty.

In 1928, Fred Griffith performed the first experiment which suggested that protein was not the genetic material. His experiment was actually fairly simple. He first injected mice with a live strain of virulent (deadly) bacteria, and not to anyone's surprise, all of those mice died. Then, he killed the virulent bacteria cells by heating them. Mice injected with these heat-killed virulent bacteria did not die. In another set of mice, Griffith injected a live *non*-virulent strain of bacteria, and these mice did not die, the result which Griffith expected.

The surprise came when Griffith injected a group of mice with both live non-virulent bacteria and heat-killed virulent bacteria. In that group, some of the mice died. When Griffith examined those mice, he found live virulent bacteria in their blood. Griffith drew the conclusion that the genetic information in the heat-killed virulent bacteria survived the heating process and was somehow incorporated into the genetic material of the non-virulent strain to cause them to become virulent. But Griffith knew that heat denatures protein, so he suggested that the genetic material must be something else. However, his results did not specifically point to DNA as a possibility.

Avery–MacLeod–McCarty experiment

Oswald Avery followed up on Griffith's experiment in the following decade. Like Griffith, Avery first used heat to kill virulent bacteria. He then extracted RNA (ribonucleic acid), DNA, carbohydrates, lipids, and proteins from these dead cells, all of which were considered to be possible candidates for the carriers of genetic information. Next, he added each type of molecule to a culture of live non-virulent bacteria to determine which was responsible for changing them into virulent bacteria as Griffith had observed. Only the non-virulent cells which were given DNA from the dead virulent strain became virulent, so Avery concluded that DNA must be the genetic material.

The Avery–MacLeod–McCarty experiment was an experimental demonstration, reported in 1944 by Oswald Avery, Colin MacLeod, and Maclyn McCarty, that DNA is the substance that causes bacterial transformation. It was the culmination of research in the 1930s and early 1940s at the Rockefeller Institute for Medical Research to purify and characterize the "transforming principle" responsible for the transformation phenomenon first described in Griffith's experiment of 1928: killed *Streptococcus pneumoniae* of the virulent strain type III-S, when injected along with living but non-virulent type II-R pneumococci, resulted in a deadly infection of type III-S pneumococci. In their paper "*Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types: Induction of Transformation by a Deoxyribonucleic Acid Fraction Isolated from Pneumococcus Type III*", published in the February 1944

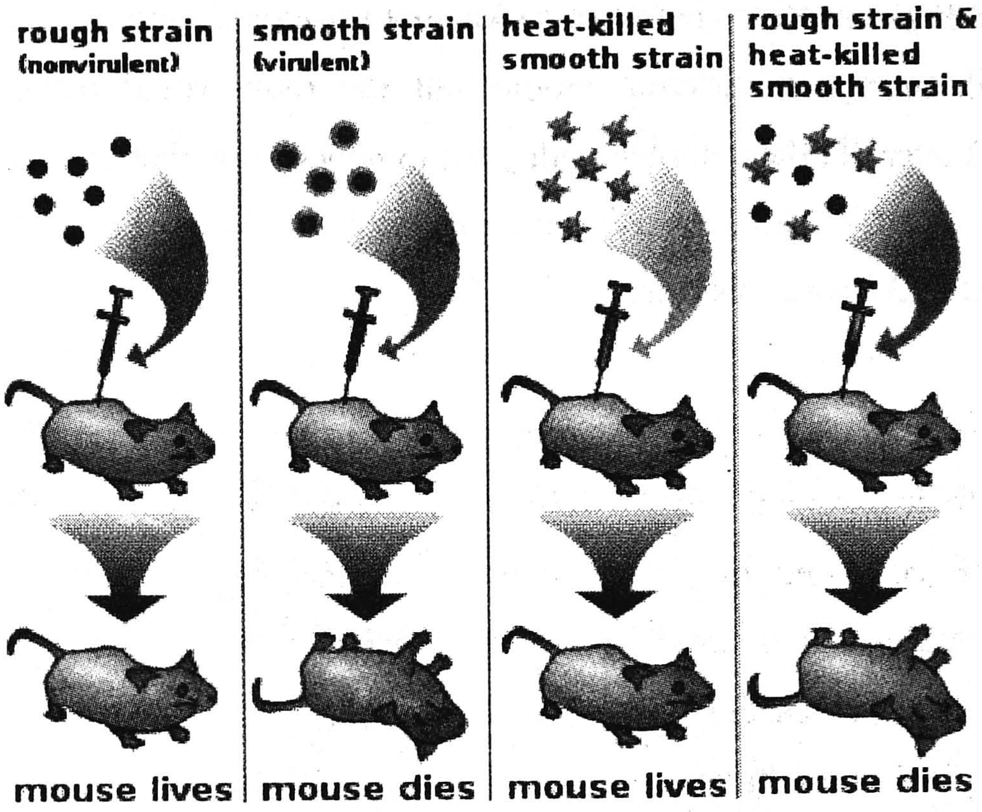
issue of the *Journal of Experimental Medicine*, Avery and his colleagues suggest that DNA, rather than protein as widely believed at the time, may be the hereditary material of bacteria, and could be analogous to genes and/or viruses in higher organisms.

Experimental work

Pneumococcus is characterized by *smooth* colonies and has a polysaccharide capsule that induces antibody formation; the different types are classified according to their immunological specificity. The purification procedure Avery et al. undertook consisted of first killing the bacteria with heat and extracting the saline-soluble components. Next, the protein was precipitated out using chloroform and the polysaccharide capsules were hydrolyzed with an enzyme. An immunological precipitation caused by type-specific antibodies was used to verify the complete destruction of the capsules. Then, the active portion was precipitated out by alcohol fractionation, resulting in fibrous strands that could be removed with a stirring rod.

Chemical analysis showed that the proportions of carbon, hydrogen, nitrogen, and phosphorus in this active portion were consistent with the chemical composition of DNA. To show that it was DNA rather than some small amount of RNA, protein, or some other cell component that was responsible for transformation, Avery and his colleagues used a number of biochemical tests. They found that trypsin, chymotrypsin and ribonuclease (enzymes that break apart proteins or RNA) did not affect it, but an enzyme preparation of "deoxyribonucleodepolymerase" (a crude preparation, obtainable from a number of animal sources, that could break down DNA) destroyed the extract's transforming power.

Follow-up work in response to criticism and challenges included the purification and crystallization, by Moses Kunitz in 1948, of a DNA depolymerase (deoxyribonuclease I), and precise work by Rollin Hotchkiss showing that virtually all the detected nitrogen in the purified DNA came from glycine, a breakdown product of the nucleotide base adenine, and that undetected protein contamination was at most 0.02% by Hotchkiss's estimation.

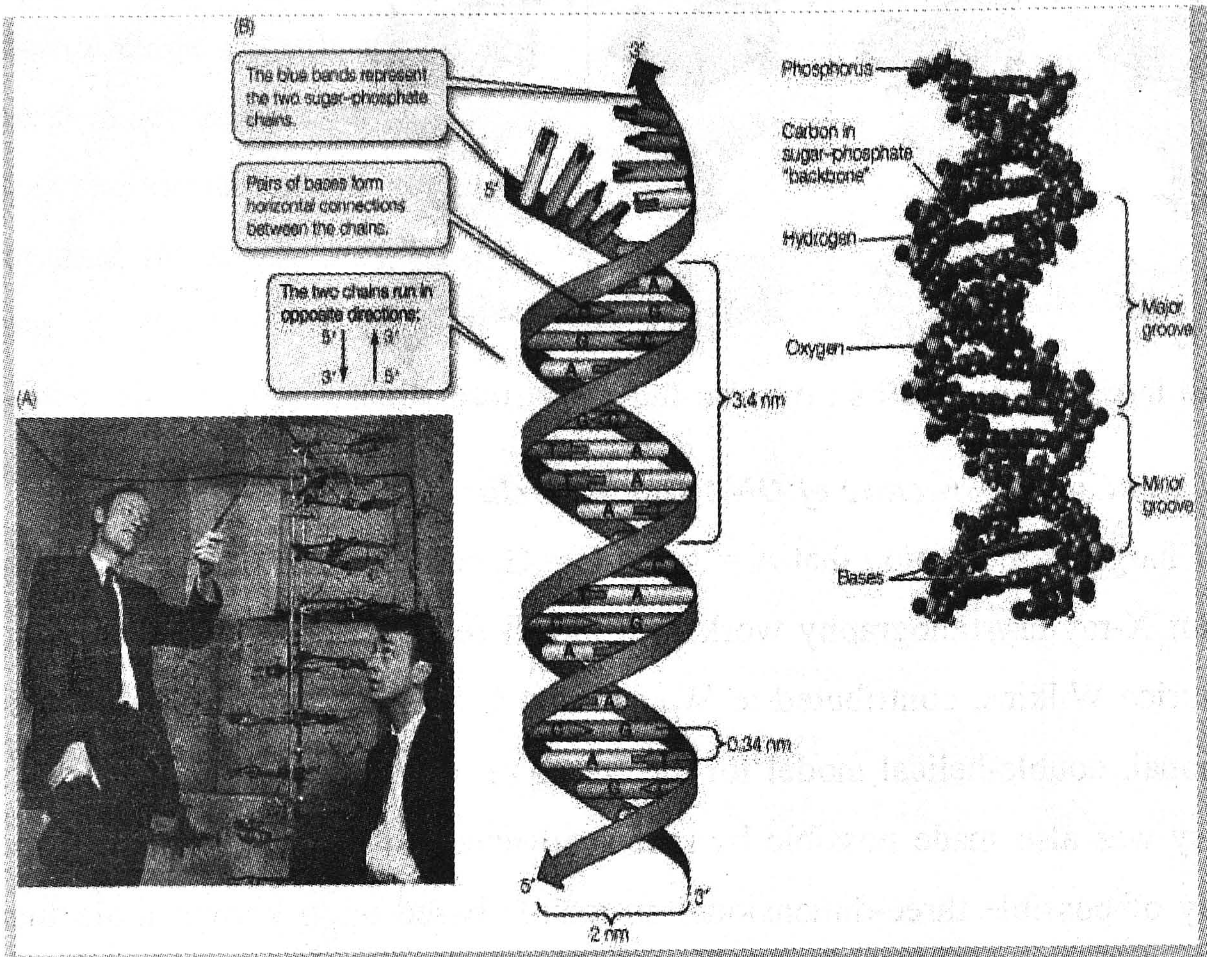


Watson and Crick's structure of DNA Double Helix

Chargaff's realization that A = T and C = G, combined with some crucially important X-ray crystallography work by English researchers Rosalind Franklin and Maurice Wilkins, contributed to Watson and Crick's derivation of the three-dimensional, double-helical model for the structure of DNA. Watson and Crick's discovery was also made possible by recent advances in model building, or the assembly of possible three-dimensional structures based upon known molecular distances and bond angles, a technique advanced by American biochemist Linus Pauling. In fact, Watson and Crick were worried that they would be "scooped" by Pauling, who proposed a different model for the three-dimensional structure of DNA just months before they did. In the end, however, Pauling's prediction was incorrect.

Using cardboard cutouts representing the individual chemical components of the four bases and other nucleotide subunits, Watson and Crick shifted molecules around on their desktops, as though putting together a puzzle. They were misled for a while by an erroneous understanding of how the different elements in thymine and guanine (specifically, the carbon, nitrogen, hydrogen, and oxygen rings) were configured. Only upon the suggestion of American

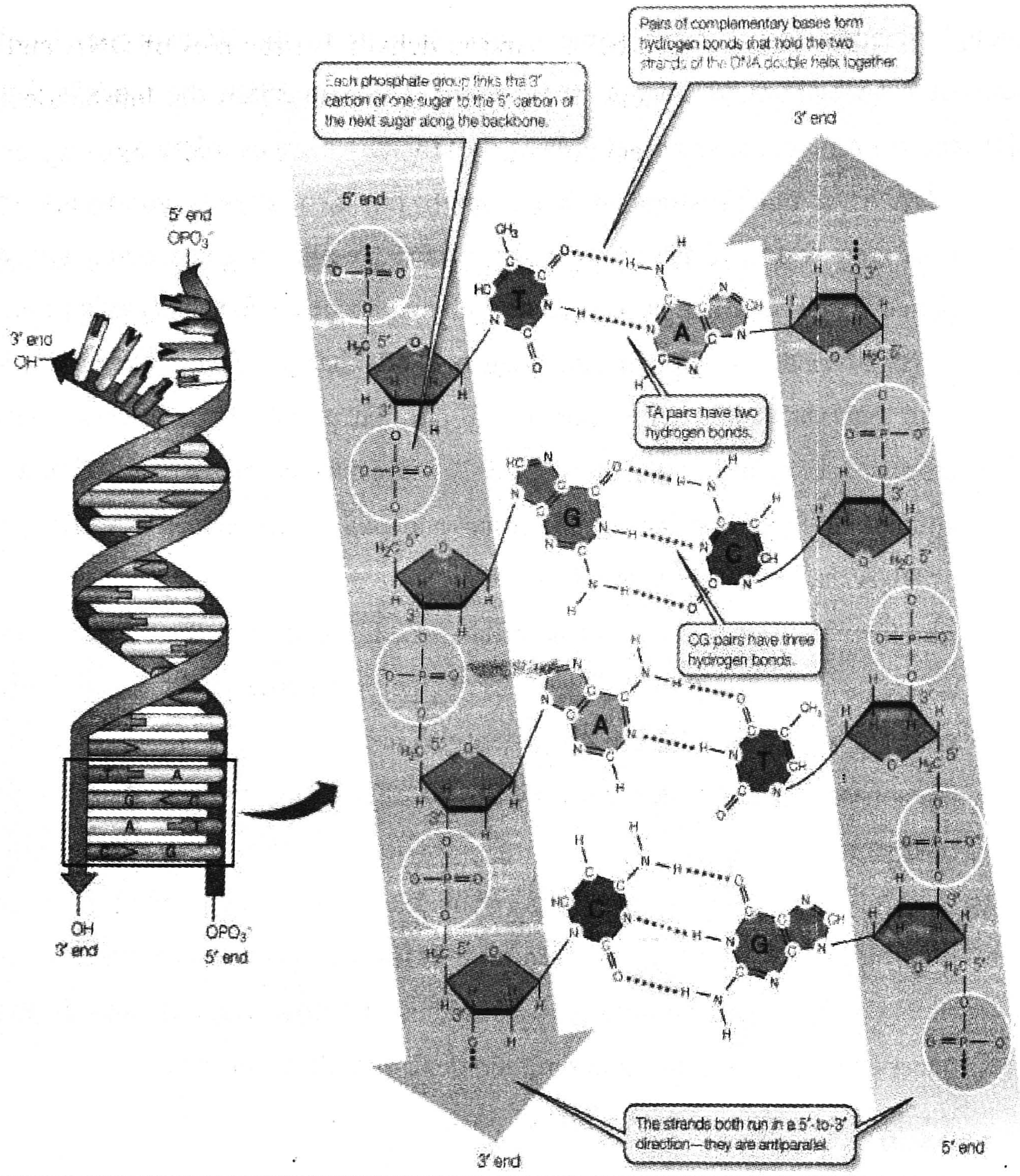
scientist Jerry Donohue did Watson decide to make new cardboard cutouts of the two bases, to see if perhaps a different atomic configuration would make a difference. It did. Not only did the complementary bases now fit together perfectly (i.e., A with T and C with G), with each pair held together by hydrogen bonds, but the structure also reflected Chargaff's rule.



DNA is a double helix.
(A) Francis Crick (left) and James Watson (right) proposed that the DNA molecule has a double-helical structure. (B) Biochemists can now pinpoint the position of every atom in a DNA molecule. To see that the essential features of the original Watson-Crick model have been verified, follow with your eyes the double-helical chains of sugar-phosphate groups and note the horizontal rungs of the bases.

Although scientists have made some minor changes to the Watson and Crick model, or have elaborated upon it, since its inception in 1953, the model's four major features remain the same yet today. These features are as follows:

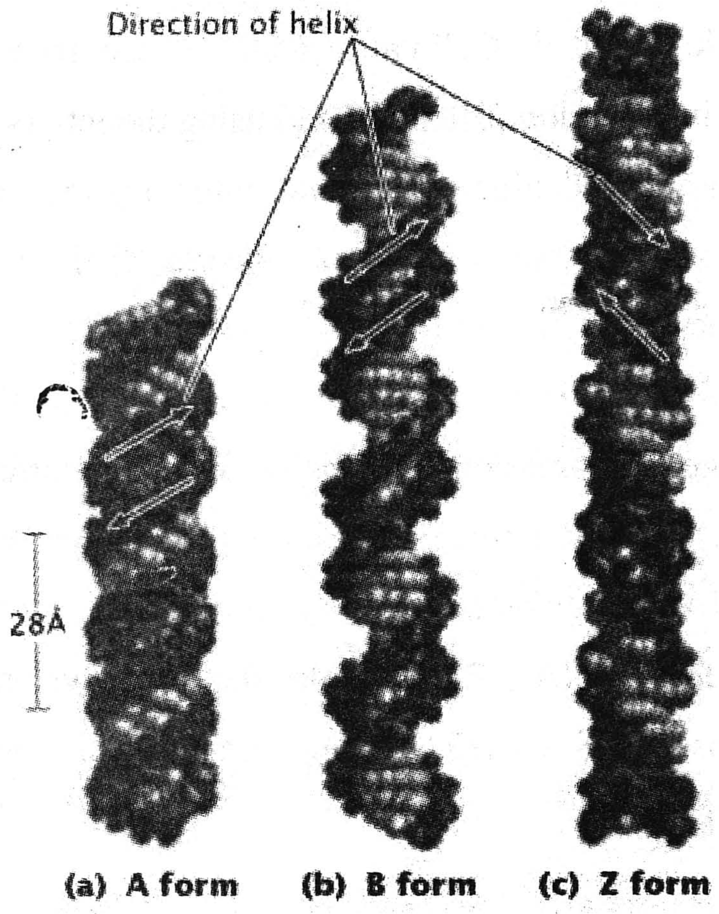
- DNA is a double-stranded helix, with the two strands connected by hydrogen bonds. A bases are always paired with Ts, and Cs are always paired with Gs, which is consistent with and accounts for Chargaff's rule.
- Most DNA double helices are right-handed; that is, if you were to hold your right hand out, with your thumb pointed up and your fingers curled around your thumb, your thumb would represent the axis of the helix and your fingers would represent the sugar-phosphate backbone. Only one type of DNA, called Z-DNA, is left-handed.
- The DNA double helix is anti-parallel, which means that the 5' end of one strand is paired with the 3' end of its complementary strand (and vice versa). As shown in Figure 4, nucleotides are linked to each other by their phosphate groups, which bind the 3' end of one sugar to the 5' end of the next sugar.
- Not only are the DNA base pairs connected via hydrogen bonding, but the outer edges of the nitrogen-containing bases are exposed and available for potential hydrogen bonding as well. These hydrogen bonds provide easy access to the DNA for other molecules, including the proteins that play vital roles in the replication and expression of DNA.



Base pairing in DNA is complementary.
The purines (A and G) pair with the pyrimidines (T and C, respectively) to form equal-sized base pairs resembling rungs on a ladder (the sugar-phosphate backbones). The ladder twists into a double-helical structure.

One of the ways that scientists have elaborated on Watson and Crick's model is through the identification of three different conformations of the DNA double helix. In other words, the precise geometries and dimensions of the double helix can vary. The most common conformation in most living cells (which is the one depicted in most diagrams of the double helix, and the one proposed by Watson and Crick) is known as B-DNA. There are also two other confirmations: A-DNA,

a shorter and wider form that has been found in dehydrated samples of DNA and rarely under normal physiological circumstances; and Z-DNA, a left-handed confirmation. Z-DNA is a transient form of DNA, only occasionally existing in response to certain types of biological activity. Z-DNA was first discovered in 1979, but its existence was largely ignored until recently. Scientists have since discovered that certain proteins bind very strongly to Z-DNA, suggesting that Z-DNA plays an important biological role in protection against viral disease.



DNA can assume several different secondary structures. These structures depend on the base sequence of the DNA and the conditions under which it is placed.

Summary

Watson and Crick were not the discoverers of DNA, but rather the first scientists to formulate an accurate description of this molecule's complex, double-helical structure. Moreover, Watson and Crick's work was directly dependent on the research of numerous scientists before them, including Friedrich Miescher, Phoebus Levene, and Erwin Chargaff. Thanks to researchers such as these, we now know a great deal about genetic structure, and we continue to make great

strides in understanding the human genome and the importance of DNA to life and health.

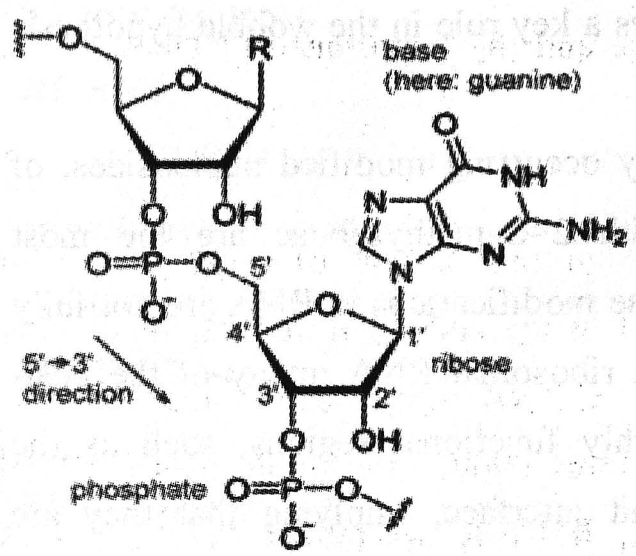
RIBONUCLEIC ACID (RNA)

Ribonucleic acid (RNA) is a ubiquitous family of large biological molecules that perform multiple vital roles in the coding, decoding, regulation, and expression of genes. Together with DNA, RNA comprises the nucleic acids, which, along with proteins, constitute the three major macromolecules essential for all known forms of life. Like DNA, RNA is assembled as a chain of nucleotides, but is usually single-stranded. Cellular organisms use messenger RNA (mRNA) to convey genetic information (often notated using the letters G, A, U, and C for the nucleotides guanine, adenine, uracil and cytosine) that directs synthesis of specific proteins, while many viruses encode their genetic information using an RNA genome.

Some RNA molecules play an active role within cells by catalyzing biological reactions, controlling gene expression, or sensing and communicating responses to cellular signals. One of these active processes is protein synthesis, a universal function whereby mRNA molecules direct the assembly of proteins on ribosomes. This process uses transfer RNA (tRNA) molecules to deliver amino acids to the ribosome, where ribosomal RNA (rRNA) links amino acids together to form proteins.

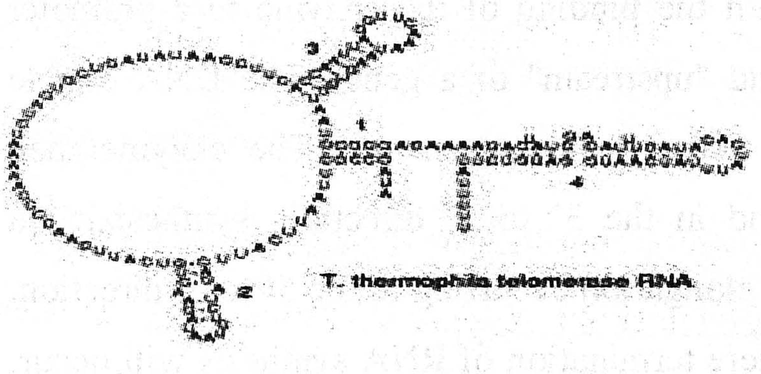
Structure

Each nucleotide in RNA contains a ribose sugar, with carbons numbered 1' through 5'. A base is attached to the 1' position, in general, adenine (A), cytosine (C), guanine (G), or uracil (U). Adenine and guanine are purines, cytosine, and uracil are pyrimidines. A phosphate group is attached to the 3' position of one ribose and the 5' position of the next. The phosphate groups have a negative charge each at physiological pH, making RNA a charged molecule (polyanion). The bases may form hydrogen bonds between cytosine and guanine, between adenine and uracil and between guanine and uracil. However, other interactions are possible, such as a group of adenine bases binding to each other in a bulge, or the GNRA tetraloop that has a guanine–adenine base-pair.



Chemical structure of RNA

An important structural feature of RNA that distinguishes it from DNA is the presence of a hydroxyl group at the 2' position of the ribose sugar. The presence of this functional group causes the helix to adopt the A-form geometry rather than the B-form most commonly observed in DNA. This results in a very deep and narrow major groove and a shallow and wide minor groove. A second consequence of the presence of the 2'-hydroxyl group is that in conformationally flexible regions of an RNA molecule (that is, not involved in formation of a double helix), it can chemically attack the adjacent phosphodiester bond to cleave the backbone.



Secondary structure of a telomerase RNA.

RNA is transcribed with only four bases (adenine, cytosine, guanine and uracil), but these bases and attached sugars can be modified in numerous ways as the RNAs mature. Pseudouridine (Ψ), in which the linkage between uracil and ribose is changed from a C–N bond to a C–C bond, and ribothymidine (T) are found in various places (the most notable ones being in the TΨC loop of tRNA). Another notable modified base is hypoxanthine, a deaminated adenine base whose

nucleoside is called inosine (I). Inosine plays a key role in the wobble hypothesis of the genetic code.

There are nearly 100 other naturally occurring modified nucleosides, of which pseudouridine and nucleosides with 2'-O-methylribose are the most common. The specific roles of many of these modifications in RNA are not fully understood. However, it is notable that, in ribosomal RNA, many of the post-transcriptional modifications occur in highly functional regions, such as the peptidyl transferase center and the subunit interface, implying that they are important for normal function.

The functional form of single-stranded RNA molecules, just like proteins, frequently requires a specific tertiary structure. The scaffold for this structure is provided by secondary structural elements that are hydrogen bonds within the molecule. This leads to several recognizable "domains" of secondary structure like hairpin loops, bulges, and internal loops. Since RNA is charged, metal ions such as Mg^{2+} are needed to stabilise many secondary and tertiary structures.

Synthesis

Synthesis of RNA is usually catalyzed by an enzyme—RNA polymerase—using DNA as a template, a process known as transcription. Initiation of transcription begins with the binding of the enzyme to a promoter sequence in the DNA (usually found "upstream" of a gene). The DNA double helix is unwound by the helicase activity of the enzyme. The enzyme then progresses along the template strand in the 3' to 5' direction, synthesizing a complementary RNA molecule with elongation occurring in the 5' to 3' direction. The DNA sequence also dictates where termination of RNA synthesis will occur. RNAs are often modified by enzymes after transcription. For example, a poly(A) tail and a 5' cap are added to eukaryotic pre-mRNA and introns are removed by the spliceosome. There are also a number of RNA-dependent RNA polymerases that use RNA as their template for synthesis of a new strand of RNA. For instance, a number of RNA viruses (such as poliovirus) use this type of enzyme to replicate their genetic material. Also, RNA-dependent RNA polymerase is part of the RNA interference pathway in many organisms.

TYPES OF RNA

RI RNA

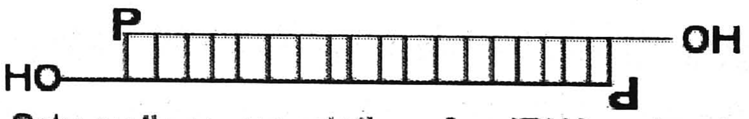
RI-RNA is a complex molecule composed of complementary negative strands and nascent positive single-stranded RNA molecules, whose synthesis is indispensable for the progress of the replicative cycle.

si RNA

Small interfering RNA (siRNA), sometimes known as short interfering RNA or silencing RNA, is a class of double-stranded RNA molecules, 20-25 base pairs in length. siRNA plays many roles, but its most notable is in the RNA interference (RNAi) pathway, where it interferes with the expression of specific genes with complementary nucleotide sequence. siRNA also acts in RNAi-related pathways, e.g., as an antiviral mechanism or in shaping the chromatin structure of a genome. The complexity of these pathways is only now being elucidated.

siRNAs and their role in post-transcriptional gene silencing (PTGS) in plants were first discovered by David Baulcombe's group at the Sainsbury Laboratory in Norwich, England and reported in *Science* in 1999. Thomas Tuschl and colleagues soon reported in *Nature* that synthetic siRNAs could induce RNAi in mammalian cells. This discovery led to a surge in interest in harnessing RNAi for biomedical research and drug development.

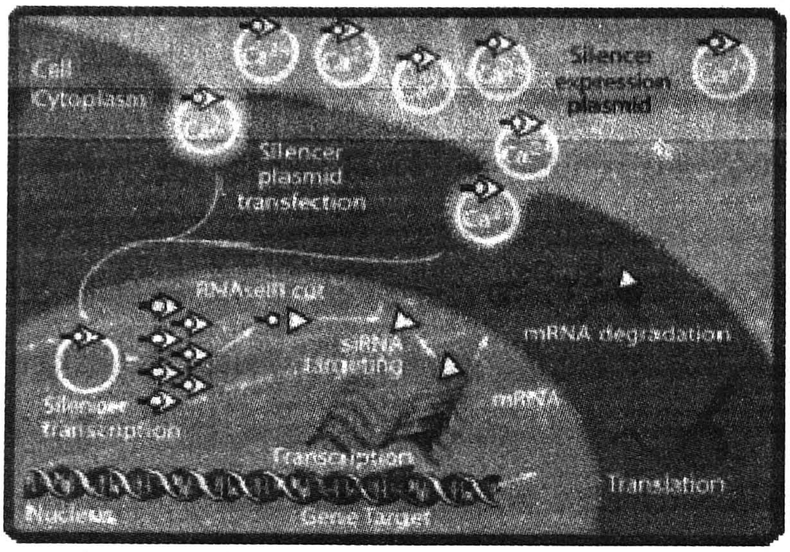
siRNAs have a well-defined structure: a short (usually 21-bp) double-stranded RNA (dsRNA) with phosphorylated 5' ends and hydroxylated 3' ends with two overhanging nucleotides:



Schematic representation of a siRNA molecule: a ~19-21 basepair RNA core duplex that is followed by a 2 nucelotide 3' overhang on each strand. OH: 3' hydroxyl; P: 5' Phosphate

The Dicer enzyme catalyzes production of siRNAs from long dsRNAs and small hairpin RNAs.[3] siRNAs can also be introduced by transfection. Since in principle any gene can be knocked down by a synthetic siRNA with a

complementary sequence, siRNAs are an important tool for validating gene function and drug targeting in the post-genomic era.



DNA REPLICATION

DNA replication is a biological process that occurs in all living organisms and copies their DNA. DNA replication during interphase is the basis for biological inheritance. The process of DNA replication starts when one double-stranded DNA molecule produces two identical copies of the molecule. Each strand of the original double-stranded DNA molecule serves as template for the production of the complementary strand, a process referred to as semiconservative replication. Cellular proofreading and error-checking mechanisms ensure near perfect fidelity for DNA replication.

In a cell, DNA replication begins at specific locations, or origin of replication, in the genome. Unwinding of DNA at the origin, and synthesis of new strands, forms a replication fork. A number of proteins are associated with the fork and assist in the initiation and continuation of DNA synthesis. Most prominently, DNA polymerase synthesizes the new DNA by adding matching nucleotides to the template strand.

DNA replication can also be performed *in vitro* (artificially, outside a cell). DNA polymerases isolated from cells and artificial DNA primers can be used to initiate DNA synthesis at known sequences in a template DNA molecule. The polymerase chain reaction (PCR), a common laboratory technique, cyclically apply such artificial synthesis to amplify a specific target DNA fragment from a pool of DNA.

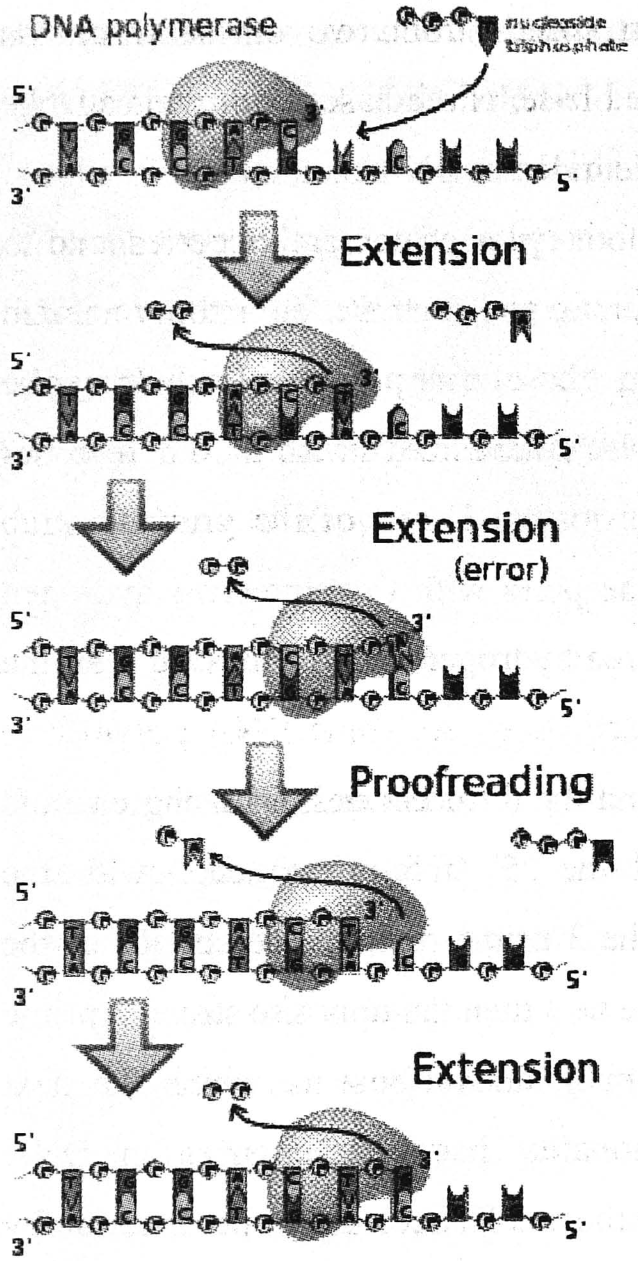
Background on DNA structure

DNA usually exists as a double-stranded structure, with both strands coiled together to form the characteristic double-helix. Each single strand of DNA is a chain of four types of nucleotides. Nucleotides in DNA contain a deoxyribose sugar, a phosphate, and a nucleobase. The four types of nucleotide correspond to the four nucleobases adenine, cytosine, guanine, and thymine, commonly notated as A,C, G and T. These nucleotides form phosphodiester bonds, creating the phosphate-deoxyribose backbone of the DNA double helix with the nucleobases pointing inward. Nucleotides (bases) are matched between strands through hydrogen bonds to form base pairs. Adenine pairs with thymine (two hydrogen bonds), and cytosine pairs with guanine (three hydrogen bonds) because a purine must pair with a pyrimidine.

DNA strands have a directionality, and the different ends of a single strand are called the "3' (three-prime) end" and the "5' (five-prime) end" with the direction of the naming going 5 prime to the 3 prime region. The strands of the helix are anti-parallel with one being 5 prime to 3 then the opposite strand 3 prime to 5. These terms refer to the carbon atom in deoxyribose to which the next phosphate in the chain attaches. Directionality has consequences in DNA synthesis, because DNA polymerase can synthesize DNA in only one direction by adding nucleotides to the 3' end of a DNA strand.

The pairing of bases in DNA through hydrogen bonding means that the information contained within each strand is redundant. The nucleotides on a single strand can be used to reconstruct nucleotides on a newly synthesized partner strand.

DNA polymerase



DNA polymerases adds nucleotides to the 3' end of a strand of DNA. If a mismatch is accidentally incorporated, the polymerase is inhibited from further extension. Proofreading removes the mismatched nucleotide and extension continues.

DNA polymerases are a family of enzymes that carry out all forms of DNA replication. However, a DNA polymerase can only extend an existing DNA strand paired with a template strand; it cannot begin the synthesis of a new strand. To begin synthesis, a short fragment of DNA or RNA, called a primer, must be created and paired with the template DNA strand.

DNA polymerase then synthesizes a new strand of DNA by extending the 3' end of an existing nucleotide chain, adding new nucleotides matched to the

template strand one at a time via the creation of phosphodiester bonds. The energy for this process of DNA polymerization comes from two of the three total phosphates attached to each unincorporated base. (Free bases with their attached phosphate groups are called nucleoside triphosphates.) When a nucleotide is being added to a growing DNA strand, two of the phosphates are removed and the energy produced creates a phosphodiester bond that attaches the remaining phosphate to the growing chain. The energetics of this process also help explain the directionality of synthesis—if DNA were synthesized in the 3' to 5' direction, the energy for the process would come from the 5' end of the growing strand rather than from free nucleotides.

In general, DNA polymerases are extremely accurate, making less than one mistake for every 10^7 nucleotides added. Even so, some DNA polymerases also have proofreading ability; they can remove nucleotides from the end of a strand in order to correct mismatched bases. If the 5' nucleotide needs to be removed during proofreading, the triphosphate end is lost. Hence, the energy source that usually provides energy to add a new nucleotide is also lost.

The rate of DNA replication in a living cell was first measured as the rate of phage T4 DNA elongation in phage-infected E. coli. During the period of exponential DNA increase at 30°C, the rate was 749 nucleotides per second. The mutation rate per base pair per replication during phage T4 DNA synthesis is 1.7 per 10^{-8}. Thus DNA replication is both impressively fast and accurate.

Replication process

DNA Replication, like all biological polymerization processes, proceeds in three enzymatically catalyzed and coordinated steps: initiation, elongation and termination.

Replication origin

For a cell to divide, it must first replicate its DNA. This process is initiated at particular points in the DNA, known as "origins", which are targeted by proteins that initiate DNA synthesis. Origins contain DNA sequences recognized by replication *initiator proteins* (e.g., DnaA in *E. coli* and the Origin Recognition Complex in yeast). Sequences used by initiator proteins tend to be

"AT-rich" (rich in adenine and thymine bases), because A-T base pairs have two hydrogen bonds (rather than the three formed in a C-G pair). AT-rich sequences are easier to unzip since less energy is required to break relatively fewer hydrogen bonds. Once the origin has been located, these initiators recruit other proteins and form the pre-replication complex, which unzips, or separates, the DNA strands at the origin.

Extensions

All known DNA replication systems require a free 3' hydroxyl group before synthesis can be initiated (Important note: DNA is read in 3' to 5' direction whereas a new strand is synthesised in the 5' to 3' direction—this is often confused). Four distinct mechanisms for synthesis have been described.

1. All cellular life forms and many DNA viruses, phages and plasmids use a primase to synthesize a short RNA primer with a free 3′ OH group which is subsequently elongated by a DNA polymerase.
2. The retroelements (including retroviruses) employ a transfer RNA that primes DNA replication by providing a free 3′ OH that is used for elongation by the reverse transcriptase.
3. In the adenoviruses and the φ29 family of bacteriophages, the 3' OH group is provided by the side chain of an amino acid of the genome attached protein (the terminal protein) to which nucleotides are added by the DNA polymerase to form a new strand.
4. In the single stranded DNA viruses — a group that includes the circoviruses, the geminiviruses, the parvoviruses and others — and also the many phages and plasmids that use the rolling circle replication (RCR) mechanism, the RCR endonuclease creates a nick the genome strand (single stranded viruses) or one of the DNA strands (plasmids). The 5′ end of the nicked strand is transferred to a tyrosine residue on the nuclease and the free 3′ OH group is then used by the DNA polymerase for new strand synthesis.

The first is the best known of these mechanisms and is used by the cellular organisms. In this mechanism, once the two strands are separated, primase adds an

RNA primers to the template strands. The leading strand receives one RNA primer while the lagging strand receives several. The leading strand is extended from the primer in one motion by DNA polymerase, while the lagging strand is extended discontinuously from each primer, forming Okazaki fragments. RNase removes the primer RNA fragments, and another DNA Polymerase enters to fill the gaps. When this is complete, a single nick on the leading strand and several nicks on the lagging strand can be found. Ligase works to fill these nicks in, thus completing the newly replicated DNA molecule.

The primase used in this process differs significantly between bacteria and archaea/eukaryotes. Bacteria use a primase belonging to the DnaG protein superfamily which contains a catalytic domain of the TOPRIM fold type. The TOPRIM fold contains an α/β core with four conserved strands in a Rossmann-like topology. This structure is also found in the catalytic domains of topoisomerase Ia, topoisomerase II, the OLD-family nucleases and DNA repair proteins related to the RecR protein.

The primase used by archaea and eukaryotes in contrast contains a highly derived version of the RNA recognition motif (RRM). This primase is structurally similar to many viral RNA dependent RNA polymerases, reverse transcriptases, cyclic nucleotide generating cyclases and DNA polymerases of the A/B/Y families that are involved in DNA replication and repair. All these proteins share a catalytic mechanism of di-metal-ion-mediated nucleotide transfer, whereby two acidic residues located at the end of the first strand and between the second and third strands of the RRM-like unit respectively, chelate two divalent cations.

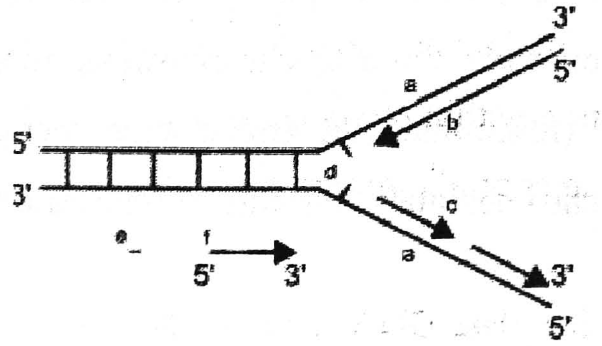
As DNA synthesis continues, the original DNA strands continue to unwind on each side of the bubble, forming a replication fork with two prongs. In bacteria, which have a single origin of replication on their circular chromosome, this process eventually creates a "theta structure" (resembling the Greek letter theta: θ). In contrast, eukaryotes have longer linear chromosomes and initiate replication at multiple origins within these.

DNA replication proteins

List of major DNA replication enzymes in the Replisome

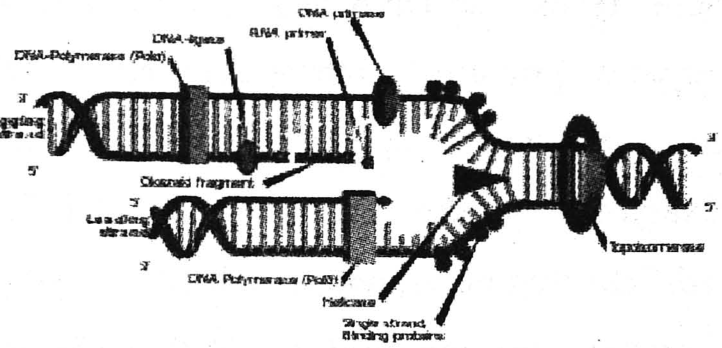
Enzyme	Function in DNA replication
DNA Helicase	Also known as helix destabilizing enzyme. Unwinds the DNA double helix at the Replication Fork.
DNA Polymerase	Builds a new duplex DNA strand by adding nucleotides in the 5' to 3' direction. Also performs proof-reading and error correction.
DNA clamp	A protein which prevents DNA polymerase III from dissociating from the DNA parent strand.
Single-Strand Binding (SSB) Proteins	Bind to ssDNA and prevent the DNA double helix from re-annealing after DNA helicase unwinds it thus maintaining the strand separation.
Topoisomerase	Relaxes the DNA from its super-coiled nature.
DNA Gyrase	Relieves strain of unwinding by DNA helicase; this is a specific type of topisomerase
DNA Ligase	Re-anneals the semi-conservative strands and joins Okazaki Fragments of the lagging strand.
Primase	Provides a starting point of RNA (or DNA) for DNA polymerase to begin synthesis of the new DNA strand.
Telomerase	Lengthens telomeric DNA by adding repetitive nucleotide sequences to the ends of **eukaryotic chromosomes**.

Replication fork



Scheme of the replication fork.

a: template, b: leading strand, c: lagging strand, d: replication fork, e: primer, f: Okazaki fragments



Many enzymes are involved in the DNA replication fork.

The replication fork is a structure that forms within the nucleus during DNA replication. It is created by helicases, which break the hydrogen bonds holding the two DNA strands together. The resulting structure has two branching "prongs", each one made up of a single strand of DNA. These two strands serve as the template for the leading and lagging strands, which will be created as DNA polymerase matches complementary nucleotides to the templates; the templates may be properly referred to as the leading strand template and the lagging strand templates

Leading strand

The leading strand is the template strand of the DNA double helix so that the replication fork moves along it in the 3' to 5' direction. This allows the newly synthesized strand complementary to the original strand to be synthesized 5' to 3' in the same direction as the movement of the replication fork.

On the leading strand, a polymerase "reads" the DNA and adds nucleotides to it continuously. This polymerase is DNA polymerase III (DNA Pol III) in prokaryotes and presumably Pol ε in yeasts. In human cells the leading and lagging strands are synthesized by Pol α and Pol δ within the nucleus and Pol γ in the mitochondria. Pol ε can substitute for Pol δ in special circumstances.

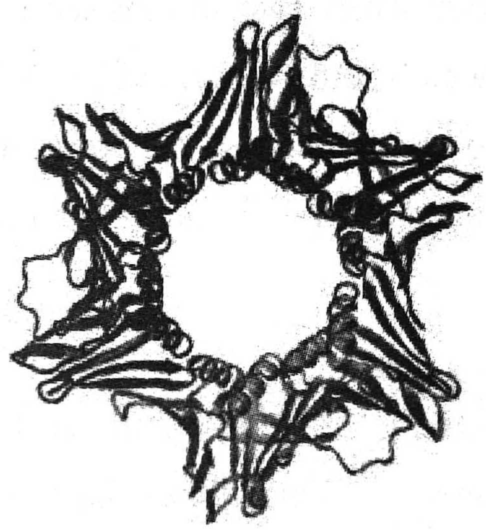
Lagging strand

The lagging strand is the strand of the template DNA double helix that is oriented so that the replication fork moves along it in a 5' to 3' manner. Because of

its orientation, opposite to the working orientation of DNA polymerase III, which moves on a template in a 3' to 5' manner, replication of the lagging strand is more complicated than that of the leading strand.

On the lagging strand, primase "reads" the DNA and adds RNA to it in short, separated segments. In eukaryotes, primase is intrinsic to Pol α. DNA polymerase III or Pol δ lengthens the primed segments, forming Okazaki fragments. Primer removal in eukaryotes is also performed by Pol δ. In prokaryotes, DNA polymerase I "reads" the fragments, removes the RNA using its flap endonuclease domain (RNA primers are removed by 5'-3' exonuclease activity of polymerase I, and replaces the RNA nucleotides with DNA nucleotides (this is necessary because RNA and DNA use slightly different kinds of nucleotides). DNA ligase joins the fragments together.

Dynamics at the replication fork



The assembled human DNA clamp, a trimer of the protein PCNA.

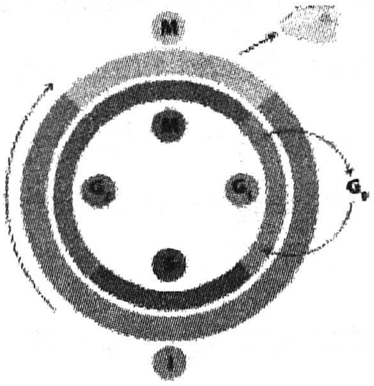
As helicase unwinds DNA at the replication fork, the DNA ahead is forced to rotate. This process results in a build-up of twists in the DNA ahead. This build-up would form a resistance that would eventually halt the progress of the replication fork. DNA Gyrase is an enzyme that temporarily breaks the strands of DNA, relieving the tension caused by unwinding the two strands of the DNA helix; DNA Gyrase achieves this by adding negative supercoils to the DNA helix.

Bare single-stranded DNA tends to fold back on itself and form secondary structures; these structures can interfere with the movement of DNA polymerase.

To prevent this, single-strand binding proteins bind to the DNA until a second strand is synthesized, preventing secondary structure formation.

Clamp proteins form a sliding clamp around DNA, helping the DNA polymerase maintain contact with its template, thereby assisting with processivity. The inner face of the clamp enables DNA to be threaded through it. Once the polymerase reaches the end of the template or detects double-stranded DNA, the sliding clamp undergoes a conformational change that releases the DNA polymerase. Clamp-loading proteins are used to initially load the clamp, recognizing the junction between template and RNA primers.

Regulation



The cell cycle of eukaryotic cells.

Eukaryotes

Within eukaryotes, DNA replication is controlled within the context of the cell cycle. As the cell grows and divides, it progresses through stages in the cell cycle; DNA replication occurs during the S phase (synthesis phase). The progress of the eukaryotic cell through the cycle is controlled by cell cycle checkpoints. Progression through checkpoints is controlled through complex interactions between various proteins, including cyclins and cyclin-dependent kinases.

The G1/S checkpoint (or restriction checkpoint) regulates whether eukaryotic cells enter the process of DNA replication and subsequent division. Cells that do not proceed through this checkpoint remain in the G0 stage and do not replicate their DNA.

Replication of chloroplast and mitochondrial genomes occurs independent of the cell cycle, through the process of D-loop replication.

Bacteria

Most bacteria do not go through a well-defined cell cycle but instead continuously copy their DNA; during rapid growth, this can result in the concurrent occurrences of multiple rounds of replication. In *E. coli*, the best-characterized bacteria, DNA replication is regulated through several mechanisms, including: the hemimethylation and sequestering of the origin sequence, the ratio of ATP to ADP, and the levels of protein DnaA. All these control the process of initiator proteins binding to the origin sequences.

Because *E. coli* methylates GATC DNA sequences, DNA synthesis results in hemimethylated sequences. This hemimethylated DNA is recognized by the protein SeqA, which binds and sequesters the origin sequence; in addition, DnaA (required for initiation of replication) binds less well to hemimethylated DNA. As a result, newly replicated origins are prevented from immediately initiating another round of DNA replication.

ATP builds up when the cell is in a rich medium, triggering DNA replication once the cell has reached a specific size. ATP competes with ADP to bind to DnaA, and the DnaA-ATP complex is able to initiate replication. A certain number of DnaA proteins are also required for DNA replication — each time the origin is copied, the number of binding sites for DnaA doubles, requiring the synthesis of more DnaA to enable another initiation of replication.

Termination

Eukaryotes initiate DNA replication at multiple points in the chromosome, so replication forks meet and terminate at many points in the chromosome; these are not known to be regulated in any particular way. Because eukaryotes have linear chromosomes, DNA replication is unable to reach the very end of the chromosomes, but ends at the telomere region of repetitive DNA close to the end. This shortens the telomere of the daughter DNA strand. This is a normal process in somatic cells. As a result, cells can only divide a certain number of times before the DNA loss prevents further division. (This is known as the Hayflick limit.) Within the germ cell line, which passes DNA to the next generation, telomerase extends the repetitive sequences of the telomere region to prevent degradation.

Telomerase can become mistakenly active in somatic cells, sometimes leading to cancer formation.

Additionally, to aid termination, the progress of the DNA replication fork must stop or be blocked. Essentially, there are two methods that organisms do this, firstly, it is to have a termination site sequence in the DNA, and secondly, it is to have a protein which binds to this sequence to physically stop DNA replication proceeding. This is named the DNA replication terminus site-binding protein or in other words, Ter protein.

Because bacteria have circular chromosomes, termination of replication occurs when the two replication forks meet each other on the opposite end of the parental chromosome. *E coli* regulate this process through the use of termination sequences that, when bound by the Tus protein, enable only one direction of replication fork to pass through. As a result, the replication forks are constrained to always meet within the termination region of the chromosome.

TRANSCRIPTION

Transcription is the first step of gene expression, in which a particular segment of DNA is copied into RNA by the enzyme, RNA polymerase. Both RNA and DNA are nucleic acids, which use base pairs of nucleotides as a complementary language that can be converted back and forth from DNA to RNA by the action of the correct enzymes. During transcription, a DNA sequence is read by an RNA polymerase, which produces a complementary, antiparallel RNA strand. As opposed to DNA replication, transcription results in an RNA complement that includes uracil (U) in all instances where thymine (T) would have occurred in a DNA complement. Also unlike DNA replication where DNA is synthesised, transcription does not involve an RNA primer to initiate RNA synthesis.

Transcription proceeds in 5 or 6 steps, each moving like a wave along the DNA.

1. One or more sigma factors initiate transcription of a gene by enabling binding of RNA polymerase to promoter DNA.
2. Helicase enzymes move a transcription bubble, like the slider of a zipper, which splits the double helix DNA molecule into two strands of unpaired

DNA nucleotides, by breaking the hydrogen bonds between complementary DNA nucleotides.

3. RNA polymerase adds matching RNA nucleotides that are paired with complementary DNA nucleotides of one DNA strand.
4. RNA sugar-phosphate backbone forms with assistance from RNA polymerase to form an RNA strand.
5. Hydrogen bonds of the untwisted RNA + DNA helix break, freeing the newly synthesized RNA strand.
6. If the cell has a nucleus, the RNA is further processed (addition of a 3'UTR poly-A tail and a 5'UTR cap) and exits to the cytoplasm through the nuclear pore complex.

Transcription is the first step leading to gene expression. The stretch of DNA transcribed into an RNA molecule is called a *transcription unit* and encodes at least one gene. If the gene transcribed encodes a protein, the result of transcription is messenger RNA (mRNA), which will then be used to create that protein via the process of translation. Alternatively, the transcribed gene may encode for either non-coding RNA genes (such as microRNA, lincRNA, etc.) or ribosomal RNA (rRNA) or transfer RNA (tRNA), other components of the protein-assembly process, or other ribozymes.

A DNA transcription unit encoding for a protein contains not only the sequence that will eventually be directly translated into the protein (the *coding sequence*) but also *regulatory sequences* that direct and regulate the synthesis of that protein. The regulatory sequence before (i.e., upstream from) the coding sequence is called the five prime untranslated region (5'UTR), and the sequence following (downstream from) the coding sequence is called the three prime untranslated region (3'UTR).

Transcription has some proofreading mechanisms, but they are fewer and less effective than the controls for copying DNA; therefore, transcription has a lower copying fidelity than DNA replication.

As in DNA replication, DNA is read from 3'UTR → 5'UTR during transcription. Meanwhile, the complementary RNA is created from the 5'UTR →

3'UTR direction. This means its 5' end is created first in base pairing. Although DNA is arranged as two antiparallel strands in a double helix, only one of the two DNA strands, called the template strand, is used for transcription. This is because RNA is only single-stranded, as opposed to double-stranded DNA. The other DNA strand is called the coding (lagging) strand, because its sequence is the same as the newly created RNA transcript (except for the substitution of uracil for thymine). The use of only the 3'UTR → 5'UTR strand eliminates the need for the Okazaki fragments seen in DNA replication.

Transcription is divided into five stages: *pre-initiation*, *initiation*, *promoter clearance*, *elongation* and *termination*.

Major steps

Pre-initiation

In eukaryotes, RNA polymerase, and therefore the initiation of transcription, requires the presence of a core promoter sequence in the DNA. Promoters are regions of DNA that promote transcription and, in eukaryotes, are found at -30, -75, and -90 base pairs upstream from the transcription start site (abbreviated to TSS). Core promoters are sequences within the promoter that are essential for transcription initiation. RNA polymerase is able to bind to core promoters in the presence of various specific transcription factors.

The most characterized type of core promoter in eukaryotes is a short DNA sequence known as a TATA box, found 25-30 base pairs upstream from the TSS. The TATA box, as a core promoter, is the binding site for a transcription factor known as TATA-binding protein (TBP), which is itself a subunit of another transcription factor, called Transcription Factor II D (TFIID). After TFIID binds to the TATA box via the TBP, five more transcription factors and RNA polymerase combine around the TATA box in a series of stages to form a preinitiation complex. One transcription factor, Transcription factor II H, has two components with helicase activity and so is involved in the separating of opposing strands of double-stranded DNA to form the initial transcription bubble. However, only a low, or basal, rate of transcription is driven by the preinitiation complex alone. Other proteins known as activators and repressors, along with any

associated coactivators or corepressors, are responsible for modulating transcription rate.

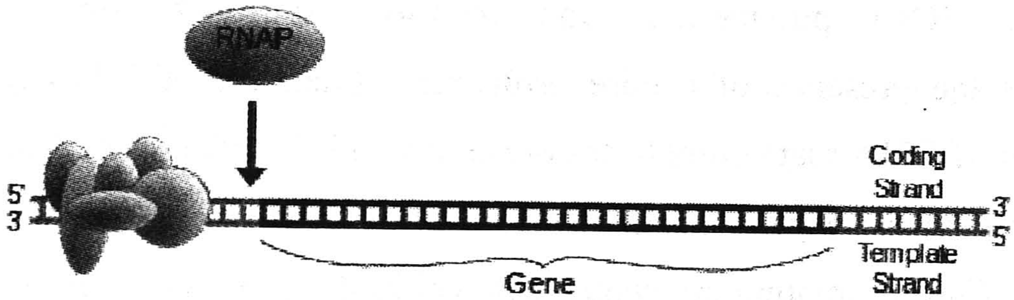
Thus, preinitiation complex contains:

Core Promoter Sequence

1. Transcription Factors
2. RNA Polymerase
3. Activators and Repressors.

The transcription preinitiation in archaea is, in essence, homologous to that of eukaryotes, but is much less complex. The archaeal preinitiation complex assembles at a TATA-box binding site; however, in archaea, this complex is composed of only RNA polymerase II, TBP, and TFB (the archaeal homologue of eukaryotic transcription factor II B (TFIIB)).

Initiation



Simple diagram of transcription initiation. RNAP = RNA polymerase

In bacteria, transcription begins with the binding of RNA polymerase to the promoter in DNA. RNA polymerase is a core enzyme consisting of five subunits: 2 α subunits, 1 β subunit, 1 β' subunit, and 1 ω subunit. At the start of initiation, the core enzyme is associated with a sigma factor that aids in finding the appropriate -35 and -10 base pairs downstream of promoter sequences. When the sigma factor and RNA polymerase combine, they form a holoenzyme.

Transcription initiation is more complex in eukaryotes. Eukaryotic RNA polymerase does not directly recognize the core promoter sequences. Instead, a collection of proteins called transcription factors mediate the binding of RNA polymerase and the initiation of transcription. Only after certain transcription factors are attached to the promoter does the RNA polymerase bind to it. The

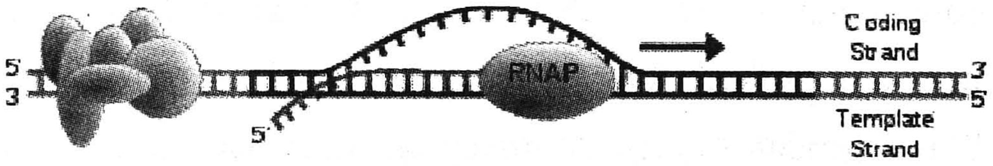
completed assembly of transcription factors and RNA polymerase bind to the promoter, forming a transcription initiation complex. Transcription in the archaea domain is similar to transcription in eukaryotes.

Promoter clearance

After the first bond is synthesized, the RNA polymerase must clear the promoter. During this time there is a tendency to release the RNA transcript and produce truncated transcripts. This is called *abortive initiation* and is common for both eukaryotes and prokaryotes. Abortive initiation continues to occur until the σ factor rearranges, resulting in the transcription elongation complex (which gives a 35 bp moving footprint). The σ factor is released before 80 nucleotides of mRNA are synthesized. Once the transcript reaches approximately 23 nucleotides, it no longer slips and elongation can occur. This, like most of the remainder of transcription, is an energy-dependent process, consuming adenosine triphosphate (ATP).

Promoter clearance coincides with phosphorylation of serine 5 on the carboxy terminal domain of RNAP II in eukaryotes, which is phosphorylated by TFIIH.

Elongation



Simple diagram of transcription elongation

One strand of the DNA, the *template strand* (or noncoding strand), is used as a template for RNA synthesis. As transcription proceeds, RNA polymerase traverses the template strand and uses base pairing complementarity with the DNA template to create an RNA copy. Although RNA polymerase traverses the template strand from 3' → 5', the coding (non-template) strand and newly formed RNA can also be used as reference points, so transcription can be described as occurring 5' → 3'. This produces an RNA molecule from 5' → 3', an exact copy of the coding strand (except that thymines are replaced with uracils, and the

nucleotides are composed of a ribose (5-carbon) sugar where DNA has deoxyribose (one less oxygen atom) in its sugar-phosphate backbone).

Unlike DNA replication, mRNA transcription can involve multiple RNA polymerases on a single DNA template and multiple rounds of transcription (amplification of particular mRNA), so many mRNA molecules can be rapidly produced from a single copy of a gene.

Elongation also involves a proofreading mechanism that can replace incorrectly incorporated bases. In eukaryotes, this may correspond with short pauses during transcription that allow appropriate RNA editing factors to bind. These pauses may be intrinsic to the RNA polymerase or due to chromatin structure.

Termination

Bacteria use two different strategies for transcription termination. 1.Rho-independent transcription 2.Rho-dependent transcription. In Rho-independent transcription termination,also called intrinsic termination, RNA transcription stops when the newly synthesized RNA molecule forms a G-C-rich hairpin loop followed by a run of Us. When the hairpin forms, the mechanical stress breaks the weak rU-dA bonds, now filling the DNA-RNA hybrid. This pulls the poly-U transcript out of the active site of the RNA polymerase, in effect, terminating transcription. In the "Rho-dependent" type of termination, a protein factor called "Rho" destabilizes the interaction between the template and the mRNA, thus releasing the newly synthesized mRNA from the elongation complex.

Transcription termination in eukaryotes is less understood but involves cleavage of the new transcript followed by template-independent addition of *A*s at its new 3' end, in a process called polyadenylation.

Prokaryotic RNA polymerase

In bacteria, the same enzyme catalyzes the synthesis of mRNA and ncRNA. RNAP is a large molecule. The core enzyme has five subunits (~400 kDa):

β': The β' subunit is the largest subunit. The β' subunit contains part of the active center responsible for RNA synthesis and contains some of the determinants for non-sequence-specific interactions with DNA and nascent RNA.

- β: The β subunit is the second-largest subunit. The β subunit contains the rest of the active center responsible for RNA synthesis and contains the rest of the determinants for non-sequence-specific interactions with DNA and nascent RNA.
- α^I and α^{II}: The α subunit is the third-largest subunit and is present in two copies per molecule of RNAP, α^I and α^{II}. Each α subunit contains two domains: αNTD (N-Terminal domain) and αCTD (C-terminal domain). αNTD contains determinants for assembly of RNAP. αCTD (C-terminal domain) contains determinants for interaction with promoter DNA, making non-sequence-non-specific interactions at most promoters and sequence-specific interactions at upstream-element-containing promoters, and contains determinants for interactions with regulatory factors.
- ω: The ω subunit is the smallest subunit. The ω subunit facilitates assembly of RNAP and stabilizes assembled RNAP.

In order to bind promoters, RNAP core associates with the transcription initiation factor sigma (σ) to form RNA polymerase holoenzyme. Sigma reduces the affinity of RNAP for nonspecific DNA while increasing specificity for promoters, allowing transcription to initiate at correct sites. The complete holoenzyme therefore has 6 subunits: $\beta'\beta\alpha^I$ and $\alpha^{II}\omega\sigma$ (~450 kDa).

GENETIC CODE

The **genetic code** is the set of rules by which information encoded within genetic material (DNA or mRNA sequences) is translated into proteins (amino acid sequences) by living cells. Biological decoding is accomplished by the ribosome, which links amino acids in an order specified by mRNA, using transfer RNA (tRNA) molecules to carry amino acids and to read the mRNA three nucleotides at a time. The genetic code is highly similar among all organisms, and can be expressed in a simple table with 64 entries.

The code defines how sequences of these nucleotide triplets, called *codons*, specify which amino acid will be added next during protein synthesis. With some exceptions, a three-nucleotide codon in a nucleic acid sequence specifies a single amino acid. Because the vast majority of genes are encoded with exactly the same code (see the RNA codon table), this particular code is often referred to as the canonical or standard genetic code, or simply *the* genetic code, though in fact some variant codes have evolved. For example, protein synthesis in human mitochondria relies on a genetic code that differs from the standard genetic code.

Not all genetic information is stored using the genetic code. All organisms' DNA contains regulatory sequences, intergenic segments, chromosomal structural areas, and other non-coding DNA that can contribute greatly to phenotype. Those elements operate under sets of rules that are distinct from the codon-to-amino acid paradigm underlying the genetic code.

The genetic code

Serious efforts to understand how proteins are encoded began after the structure of DNA was discovered by James Watson and Francis Crick, who used the experimental evidence of Maurice Wilkins and Rosalind Franklin (among others). George Gamow postulated that a three-letter code must be employed to encode the 20 standard amino acids used by living cells to build proteins. With four different nucleotides, a code of 2 nucleotides would allow for only a maximum of 4^2 or 16 amino acids. A code of 3 nucleotides could code for a maximum of 4^3 or 64 amino acids.

The Crick, Brenner et al. experiment was the first to demonstrate that codons consist of three DNA bases. Marshall Nirenberg and Heinrich J. Matthaei were the first to elucidate the nature of a codon in 1961 at the National Institutes of Health. They used a cell-free system to translate a poly-uracil RNA sequence (i.e., UUUUU...) and discovered that the polypeptide that they had synthesized consisted of only the amino acid phenylalanine. They thereby deduced that the codon UUU specified the amino acid phenylalanine. This was followed by experiments in Severo Ochoa's laboratory that demonstrated that the poly-adenine

RNA sequence (AAAAA...) coded for the polypeptide poly-lysine and that the poly-cytosine RNA sequence (CCCCC...) coded for the polypeptide poly-proline. Therefore the codon AAA specified the amino acid lysine, and the codon CCC specified the amino acid proline. Using different copolymers most of the remaining codons were then determined. Subsequent work by Har Gobind Khorana identified the rest of the genetic code. Shortly thereafter, Robert W. Holley determined the structure of transfer RNA (tRNA), the adapter molecule that facilitates the process of translating RNA into protein. This work was based upon earlier studies by Severo Ochoa, who received the Nobel prize in 1959 for his work on the enzymology of RNA synthesis.

Extending this work, Nirenberg and Philip Leder revealed the triplet nature of the genetic code and deciphered the codons of the standard genetic code. In these experiments, various combinations of mRNA were passed through a filter that contained ribosomes, the components of cells that translate RNA into protein. Unique triplets promoted the binding of specific tRNAs to the ribosome. Leder and Nirenberg were able to determine the sequences of 54 out of 64 codons in their experiments. In 1968, Khorana, Holley and Nirenberg received the Nobel Prize in Physiology or Medicine for their work.

Salient features

Sequence reading frame

A codon is defined by the initial nucleotide from which translation starts. For example, the string GGGAAACCC, if read from the first position, contains the codons GGG, AAA, and CCC; and, if read from the second position, it contains the codons GGA and AAC; if read starting from the third position, GAA and ACC. Every sequence can, thus, be read in three reading frames, each of which will produce a different amino acid sequence (in the given example, Gly-Lys-Pro, Gly-Asn, or Glu-Thr, respectively). With double-stranded DNA, there are six possible reading frames, three in the forward orientation on one strand and three reverse on the opposite strand. The actual frame in which a protein sequence is translated is defined by a start codon, usually the first AUG codon in the mRNA sequence.

Start/stop codons

Translation starts with a chain initiation codon (start codon). Unlike stop codons, the codon alone is not sufficient to begin the process. Nearby sequences (such as the Shine-Dalgarno sequence in *E. coli*) and initiation factors are also required to start translation. The most common start codon is AUG, which is read as methionine or, in bacteria, as formylmethionine. Alternative start codons (depending on the organism), include "GUG" or "UUG"; these codons normally represent valine and leucine, respectively, but, as a start codon, they are translated as methionine or formylmethionine.

The three stop codons have been given names: UAG is *amber*, UGA is *opal* (sometimes also called *umber*), and UAA is *ochre*. "Amber" was named by discoverers Richard Epstein and Charles Steinberg after their friend Harris Bernstein, whose last name means "amber" in German. The other two stop codons were named "ochre" and "opal" in order to keep the "color names" theme. Stop codons are also called "termination" or "nonsense" codons. They signal release of the nascent polypeptide from the ribosome because there is no cognate tRNA that has anticodons complementary to these stop signals, and so a release factor binds to the ribosome instead.

Effect of mutations

During the process of DNA replication, errors occasionally occur in the polymerization of the second strand. These errors, called mutations, can have an impact on the phenotype of an organism, especially if they occur within the protein coding sequence of a gene. Error rates are usually very low—1 error in every 10–100 million bases—due to the "proofreading" ability of DNA polymerases.

Missense mutations and nonsense mutations are examples of point mutations, which can cause genetic diseases such as sickle-cell disease and thalassemia respectively. Clinically important missense mutations generally change the properties of the coded amino acid residue between being basic, acidic polar or non-polar, whereas nonsense mutations result in a stop codon.

Mutations that disrupt the reading frame sequence by indels (insertions or deletions) of a non-multiple of 3 nucleotide bases are known as frameshift mutations. These mutations usually result in a completely different translation from the original, and are also very likely to cause a stop codon to be read, which truncates the creation of the protein. These mutations may impair the function of the resulting protein, and are thus rare in *in vivo* protein-coding sequences. One reason inheritance of frameshift mutations is rare is that, if the protein being translated is essential for growth under the selective pressures the organism faces, absence of a functional protein may cause death before the organism is viable. Frameshift mutations may result in severe genetic diseases such as Tay-Sachs disease.

Although most mutations that change protein sequences are harmful or neutral, some mutations have a positive effect on an organism. These mutations may enable the mutant organism to withstand particular environmental stresses better than wild-type organisms, or reproduce more quickly. In these cases a mutation will tend to become more common in a population through natural selection. Viruses that use RNA as their genetic material have rapid mutation rates, which can be an advantage, since these viruses will evolve constantly and rapidly and thus evade the defensive responses of e.g. the human immune system. In large populations of asexually reproducing organisms, for example, *E. coli*, multiple beneficial mutations may co-occur. This phenomenon is called clonal interference and causes competition among the mutations.

Degeneracy

Degeneracy is the redundancy of the genetic code. The genetic code has redundancy but no ambiguity (see the codon tables above for the full correlation). For example, although codons GAA and GAG both specify glutamic acid (redundancy), neither of them specifies any other amino acid (no ambiguity). The codons encoding one amino acid may differ in any of their three positions. For example the amino acid glutamic acid is specified by GAA and GAG codons (difference in the third position), the amino acid leucine is specified by UUA, UUG, CUU, CUC, CUA, CUG codons (difference in the first or third position),

while the amino acid serine is specified by UCA, UCG, UCC, UCU, AGU, AGC (difference in the first, second, or third position).

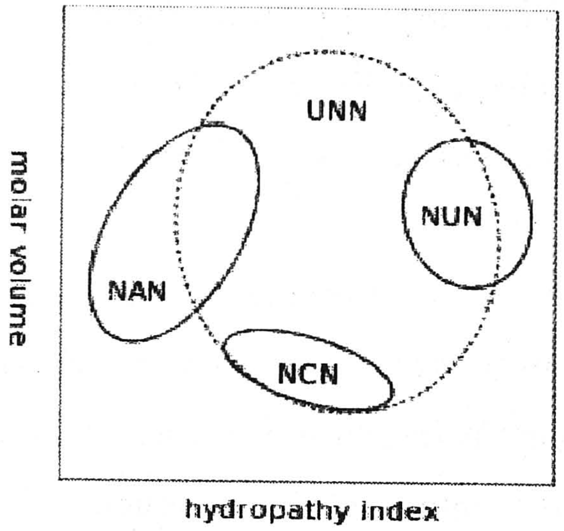
A position of a codon is said to be a fourfold degenerate site if any nucleotide at this position specifies the same amino acid. For example, the third position of the glycine codons (GGA, GGG, GGC, GGU) is a fourfold degenerate site, because all nucleotide substitutions at this site are synonymous; i.e., they do not change the amino acid. Only the third positions of some codons may be fourfold degenerate. A position of a codon is said to be a twofold degenerate site if only two of four possible nucleotides at this position specify the same amino acid. For example, the third position of the glutamic acid codons (GAA, GAG) is a twofold degenerate site. In twofold degenerate sites, the equivalent nucleotides are always either two purines (A/G) or two pyrimidines (C/U), so only transversional substitutions (purine to pyrimidine or pyrimidine to purine) in twofold degenerate sites are nonsynonymous. A position of a codon is said to be a non-degenerate site if any mutation at this position results in amino acid substitution. There is only one threefold degenerate site where changing to three of the four nucleotides may have no effect on the amino acid (depending on what it is changed to), while changing to the fourth possible nucleotide always results in an amino acid substitution. This is the third position of an isoleucine codon: AUU, AUC, or AUA all encode isoleucine, but AUG encodes methionine. In computation this position is often treated as a twofold degenerate site.

There are three amino acids encoded by six different codons: serine, leucine, and arginine. Only two amino acids are specified by a single codon. One of these is the amino-acid methionine, specified by the codon AUG, which also specifies the start of translation; the other is tryptophan, specified by the codon UGG. The degeneracy of the genetic code is what accounts for the existence of synonymous mutations.

Degeneracy results because there are more codons than encodable amino acids. For example, if there were two bases per codon, then only 16 amino acids could be coded for ($4^2=16$). Because at least 21 codes are required (20 amino acids

plus stop) and the next largest number of bases is three, then 4^3 gives 64 possible codons, meaning that some degeneracy must exist.

These properties of the genetic code make it more fault-tolerant for point mutations. For example, in theory, fourfold degenerate codons can tolerate any point mutation at the third position, although codon usage bias restricts this in practice in many organisms; twofold degenerate codons can tolerate one out of the three possible point mutations at the third position. Since transition mutations (purine to purine or pyrimidine to pyrimidine mutations) are more likely than transversion (purine to pyrimidine or vice-versa) mutations, the equivalence of purines or that of pyrimidines at twofold degenerate sites adds a further fault-tolerance.



Grouping of codons by amino acid residue molar volume and hydropathy.

A practical consequence of redundancy is that some errors in the genetic code cause only a silent mutation or an error that would not affect the protein because the hydrophilicity or hydrophobicity is maintained by equivalent substitution of amino acids; for example, a codon of NUN (where N = any nucleotide) tends to code for hydrophobic amino acids. NCN yields amino acid residues that are small in size and moderate in hydropathy; NAN encodes average size hydrophilic residues. These tendencies may result from the shared ancestry of the aminoacyl tRNA synthetases related to these codons. These variable codes for amino acids are allowed because of modified bases in the first base of the anticodon of the tRNA, and the base-pair formed is called a wobble base pair. The modified bases include inosine and the Non-Watson-Crick U-G basepair.

Transfer of information via the genetic code

The genome of an organism is inscribed in DNA, or, in the case of some viruses, RNA. The portion of the genome that codes for a protein or an RNA is called a gene. Those genes that code for proteins are composed of tri-nucleotide units called **codons**, each coding for a single amino acid. Each nucleotide sub-unit consists of a phosphate, a deoxyribose sugar, and one of the four nitrogenous nucleobases. The purine bases adenine (A) and guanine (G) are larger and consist of two aromatic rings. The pyrimidine bases cytosine (C) and thymine (T) are smaller and consist of only one aromatic ring. In the double-helix configuration, two strands of DNA are joined to each other by hydrogen bonds in an arrangement known as base pairing. These bonds almost always form between an adenine base on one strand and a thymine base on the other strand, or between a cytosine base on one strand and a guanine base on the other. This means that the number of A and T bases will be the same in a given double helix, as will the number of G and C bases. In RNA, thymine (T) is replaced by uracil (U), and the deoxyribose is substituted by ribose.

Each protein-coding gene is transcribed into a molecule of the related RNA polymer. In prokaryotes, this RNA functions as messenger RNA or mRNA; in eukaryotes, the transcript needs to be processed to produce a mature mRNA. The mRNA is, in turn, translated on a ribosome into a chain of amino acids otherwise known as a polypeptide. The process of translation requires transfer RNAs which are covalently attached to a specific amino acid, guanosine triphosphate as an energy source, and a number of translation factors. tRNAs have anticodons complementary to the codons in an mRNA and can be covalently "charged" with specific amino acids at their 3' terminal CCA ends by enzymes known as aminoacyl tRNA synthetases, which have high specificity for both their cognate amino acid and tRNA. The high specificity of these enzymes is a major reason why the fidelity of protein translation is maintained.

There are $4^3 = 64$ different codon combinations possible with a triplet codon of three nucleotides; all 64 codons are assigned to either an amino acid or a stop signal. If, for example, an RNA sequence UUUAAACCC is considered and

the reading frame starts with the first U (by convention, 5' to 3'), there are three codons, namely, UUU, AAA, and CCC, each of which specifies one amino acid. Therefore, this 9 base RNA sequence will be translated into an amino acid sequence that is three amino acids long. A given amino acid may be encoded by between one and six different codon sequences. A comparison may be made using bioinformatics tools wherein the codon is similar to a word, which is the standard data "chunk" and a nucleotide is similar to a bit, in that it is the smallest unit. This allows for powerful comparisons across species as well as within organisms.

The standard genetic code is shown in the following tables. Table 1 shows which amino acid each of the 64 codons specifies. Table 2 shows which codons specify each of the 20 standard amino acids involved in translation. These are called forward and reverse codon tables, respectively. For example, the codon "AAU" represents the amino acid asparagine, and "UGU" and "UGC" represent cysteine (standard three-letter designations, Asn and Cys, respectively).

RNA codon table

nonpolar	polar	basic	acidic	(stop codon)

Standard genetic code

1st base	2nd base U		C		A		G		3rd base
U	UUU	(Phe/F) Phenylalanine	UCU	(Ser/S) Serine	UAU	(Tyr/Y) Tyrosine	UGU	(Cys/C) Cysteine	U
	UUC		UCC		UAC		UGC		C
	UUA	(Leu/L) Leucine	UCA		UAA	Stop (*Ochre*)	UGA	Stop (*Opal*)	A
	UUG		UCG		UAG	Stop (*Amber*)	UGG	(Trp/W) Tryptophan	G
C	CUU		CCU	(Pro/P) Proline	CAU	(His/H) Histidine	CGU	(Arg/R) Arginine	U
	CUC		CCC		CAC		CGC		C
	CUA		CCA		CAA	(Gln/Q) Glutamine	CGA		A
	CUG		CCG		CAG		CGG		G
A	AUU	(Ile/I) Isoleucine	ACU	(Thr/T) Threonine	AAU	(Asn/N) Asparagine	AGU	(Ser/S) Serine	U
	AUC		ACC		AAC		AGC		C
	AUA		ACA		AAA	(Lys/K) Lysine	AGA	(Arg/R) Arginine	A
	AUG[A]	(Met/M) Methionine	ACG		AAG		AGG		G
G	GUU	(Val/V) Valine	GCU	(Ala/A) Alanine	GAU	(Asp/D) Aspartic	GGU	(Gly/G) Glycine	U

GUC		GCC		GAC	acid	GGC		C
GUA		GCA		GAA	(Glu/E) Glutamic acid	GGA		A
GUG		GCG		GAG		GGG		G

[A] The codon AUG both codes for methionine and serves as an initiation site: the first AUG in an mRNA's coding region is where translation into protein begins

Inverse table (compressed using IUPAC notation)

Amino acid	**Codons**	**Compressed**	**Amino acid**	**Codons**	**Compressed**
Ala/A	GCU, GCC, GCA, GCG	GCN	**Leu/L**	UUA, UUG, CUU, CUC, CUA, CUG	YUR, CUN
Arg/R	CGU, CGC, CGA, CGG, AGA, AGG	CGN, MGR	**Lys/K**	AAA, AAG	AAR
Asn/N	AAU, AAC	AAY	**Met/M**	AUG	
Asp/D	GAU, GAC	GAY	**Phe/F**	UUU, UUC	UUY
Cys/C	UGU, UGC	UGY	**Pro/P**	CCU, CCC, CCA, CCG	CCN
Gln/Q	CAA, CAG	CAR	**Ser/S**	UCU, UCC, UCA, UCG, AGU, AGC	UCN, AGY
Glu/E	GAA, GAG	GAR	**Thr/T**	ACU, ACC, ACA, ACG	ACN
Gly/G	GGU, GGC, GGA, GGG	GGN	**Trp/W**	UGG	
His/H	CAU, CAC	CAY	**Tyr/Y**	UAU, UAC	UAY
Ile/I	AUU, AUC, AUA	AUH	**Val/V**	GUU, GUC, GUA, GUG	GUN
START	AUG		**STOP**	UAA, UGA, UAG	UAR, URA

DNA codon table

The DNA codon table is essentially identical to that for RNA, but with U replaced by T.

Variations to the standard genetic code

While slight variations on the standard code had been predicted earlier, none were discovered until 1979, when researchers studying human mitochondrial genes discovered they used an alternative code. Many slight variants have been discovered since then, including various alternative mitochondrial codes, and small variants such as translation of the codon UGA as tryptophan in *Mycoplasma* species, and translation of CUG as a serine rather than a leucine in yeasts of the "CTG clade" (*Candida albicans* is member of this group). Because viruses must use the same genetic code as their hosts, modifications to the standard genetic code could interfere with the synthesis or functioning of viral proteins. However, some viruses (such as totiviruses) have adapted to the genetic code modification of the host. In bacteria and archaea, GUG and UUG are common start codons, but in rare cases, certain proteins may use alternative start codons not normally used by that species.

In certain proteins, non-standard amino acids are substituted for standard stop codons, depending on associated signal sequences in the messenger RNA. For example, UGA can code for selenocysteine and UAG can code for pyrrolysine. Selenocysteine is now viewed as the 21st amino acid, and pyrrolysine is viewed as the 22nd. Unlike selenocysteine, pyrrolysine encoded UAG is translated with the participation of a dedicated aminoacyl-tRNA synthetase. Both selenocysteine and pyrrolysine may be present in the same organism. Although the genetic code is normally fixed in an organism the achaeal prokaryote *Acetohalobium arabaticum* can expand its genetic code from 20 to 21 amino acids (by including pyrrolysine) under different conditions of growth.

Despite these differences, all known naturally-occurring codes are very similar to each other, and the coding mechanism is the same for all organisms: three-base codons, tRNA, ribosomes, reading the code in the same direction and translating the code three letters at a time into sequences of amino acids.

In the episode "Encyclopedia Galactica" of his TV series *Cosmos: A Personal Voyage*, Carl Sagan speculates that some intelligent extraterrestrial beings might have a genetic code based on polyaromatic sulfonyl halides instead of DNA.

Genetic code logo of the Globobulimina pseudospinescens mitochondrial genome. The logo shows the 64 codons from left to right, predicted alternatives in red (relative to the standard genetic code). Red line: stop codons. The height of each amino acid in the stack shows how often it is aligned to the codon in homologous protein domains. The stack height indicates the support for the prediction.

Predicting the genetic code

The genetic code used by a genome can be predicted by identifying the genes encoded on that genome, and comparing the codons on the DNA to the amino acids in homologous proteins in other genomes. The evolutionary conservation of protein sequences makes it possible to predict the amino acid translation for each codon as the one that is most often aligned to that codon. The program FACIL allows the automated prediction of the genetic code, searching which amino acids in homologous protein domains are most often aligned to every codon. The resulting amino acid probabilities for each codon are displayed in a genetic code logo, that also shows the support for a stop codon.

Expanded genetic code

Since 2001, 40 non-natural amino acids have been added into protein by creating a unique codon (recoding) and a corresponding transfer-RNA:aminoacyl – tRNA-synthetase pair to encode it with diverse physicochemical and biological properties in order to be used as a tool to exploring protein structure and function or to create novel or enhanced proteins.

H. Murakami and M. Sisido have extended some codons to have four and five bases. Steven A. Benner constructed a functional 65th (*in vivo*) codon.

Origin

If amino acids were randomly assigned to triplet codons, then there would be 1.5 x 10^{84} possible genetic codes to choose from. However, the genetic code used by all known forms of life is nearly universal with few minor variations. This suggests that a single evolutionary history underlies the origin of the genetic code. Many hypotheses on the evolutionary origins of the universal genetic code have been proposed.

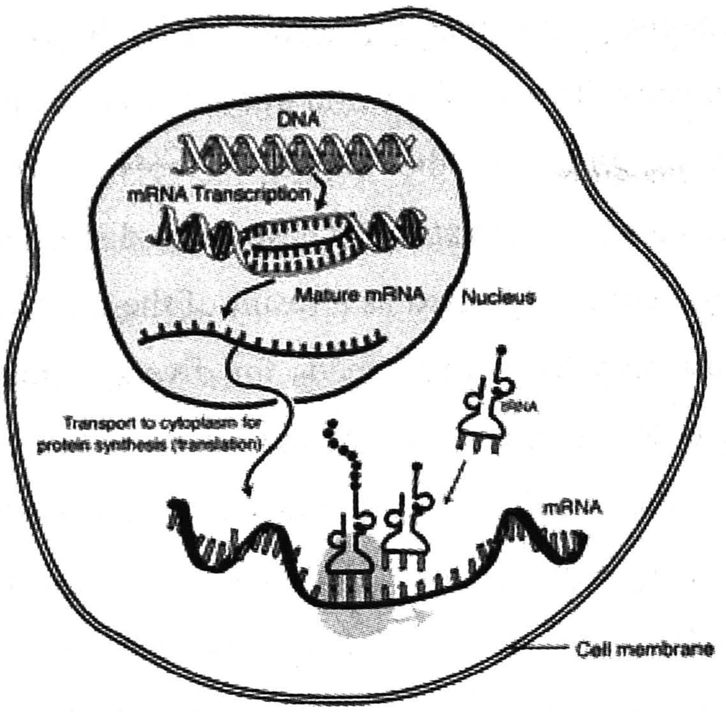
Four themes run through the many hypotheses about the evolution of the genetic code:

- **Chemical principles** govern specific RNA interaction with amino acids. Experiments with aptamers showed that some amino acids have a selective chemical affinity for the base triplets that code for them. Recent experiments show that of the 8 amino acids tested, 6 show some RNA triplet-amino acid association.
- **Biosynthetic expansion**. The standard modern genetic code grew from a simpler earlier code through a process of "biosynthetic expansion". Here the idea is that primordial life "discovered" new amino acids (for example, as by-products of metabolism) and later incorporated some of these into the machinery of genetic coding. Although much circumstantial evidence has been found to suggest that fewer different amino acids were used in the past than today, precise and detailed hypotheses about which amino acids entered the code in what order have proved far more controversial.
- **Natural selection** has led to codon assignments of the genetic code that minimize the effects of mutations. A recent hypothesis suggests that the triplet code was derived from codes that used longer than triplet codons (such as quadruplet codons). Longer than triplet decoding would have higher degree of codon redundancy and would be more error resistant than the triplet decoding. This feature could allow accurate decoding in the absence of highly complex translational machinery such as the ribosome and before cells began making ribosomes.

- **Information channels:** Information-theoretic approaches model the process of translating the genetic code into corresponding amino acids as an error-prone information channel. The inherent noise (that is, the error) in the channel poses the organism with a fundamental question: how can a genetic code be constructed to withstand the impact of noise while accurately and efficiently translating information? These "rate-distortion" models suggest that the genetic code originated as a result of the interplay of the three conflicting evolutionary forces: the needs for diverse amino-acids, for error-tolerance and for minimal cost of resources. The code emerges at a coding transition when the mapping of codons to amino-acids becomes nonrandom. The emergence of the code is governed by the topology defined by the probable errors and is related to the map coloring problem. Transfer RNA molecules appear to have evolved before modern aminoacyl-tRNA synthetases, so the latter cannot be part of the explanation of its patterns.
- There are enough data to refute the possibility that the genetic code was randomly constructed ("a frozen accident"). For example, the genetic code clusters certain amino acid assignments. Amino acids that share the same biosynthetic pathway tend to have the same first base in their codons. Amino acids with similar physical properties tend to have similar codons. A robust hypothesis for the origin of genetic code should also address or predict the following gross features of the codon table:

1. absence of codons for D-amino acids
2. secondary codon patterns for some amino acids
3. confinement of synonymous positions to third position
4. limitation to 20 amino acids instead of a number closer to 64
5. relation of stop codon patterns to amino acid coding patterns

Protein synthesis



RNA is transcribed in the nucleus; once completely processed, it is transported to the cytoplasm and translated by the ribosome (not shown).

Protein biosynthesis is the process by which biological cells generate new proteins; it is balanced by the loss of cellular proteins via degradation or export. Translation, the assembly of proteins by ribosomes, is an essential part of the biosynthetic pathway, along with generation of messenger RNA (mRNA), aminoacylation of transfer RNA (tRNA), co-translational transport, and post-translational modification. Protein biosynthesis is strictly regulated at multiple steps, and error-checking mechanisms are in place.

The cistron DNA is transcribed into a variety of RNA intermediates. The last version is used as a template in synthesis of a polypeptide chain. Protein will often be synthesized directly from genes by translating mRNA. When a protein must be available on short notice or in large quantities, a protein precursor is produced. A proprotein is an inactive protein containing one or more inhibitory peptides that can be activated when the inhibitory sequence is removed by proteolysis during posttranslational modification. A preprotein is a form that contains a signal sequence (an N-terminal signal peptide) that specifies its insertion into or through membranes, i.e., targets them for secretion. The signal

peptide is cleaved off in the endoplasmic reticulum. Preproproteins have both sequences (inhibitory and signal) still present.

In protein synthesis, a succession of tRNA molecules charged with appropriate amino acids are brought together with an mRNA molecule and matched up by base-pairing through the anti-codons of the tRNA with successive codons of the mRNA. The amino acids are then linked together to extend the growing protein chain, and the tRNAs, no longer carrying amino acids, are released. This whole complex of processes is carried out by the ribosome, formed of two main chains of RNA, called ribosomal RNA (rRNA), and more than 50 different proteins. The ribosome latches onto the end of an mRNA molecule and moves along it, capturing loaded tRNA molecules and joining together their amino acids to form a new protein chain.

Protein biosynthesis, although very similar, is different for prokaryotes and eukaryotes.

TRANSCRIPTION

In transcription an mRNA chain is generated, with one strand of the DNA double helix in the genome as a template. This strand is called the template strand. Transcription can be divided into 3 stages: initiation, elongation, and termination, each regulated by a large number of proteins such as transcription factors and coactivators that ensure that the correct gene is transcribed.

The DNA strand is read in the 3' to 5' direction and the mRNA is transcribed in the 5' to 3' direction by the RNA polymerase.

Transcription occurs in the cell nucleus, where the DNA is held. The DNA structure of the cell is made up of two helixes made up of sugar and phosphate held together by hydrogen bonds between the bases of opposite strands. The sugar and the phosphate in each strand are joined together by stronger phosphodiester covalent bonds. The DNA is "unzipped" (disruption of hydrogen bonds between different single strands) by the enzyme helicase, leaving the single nucleotide chain open to be copied. RNA polymerase reads the DNA strand from 3-prime (3') end to the 5-prime (5') end, while it synthesizes a single strand of messenger RNA in the 5'-to-3' direction. The general RNA structure is very similar to the

DNA structure, but in RNA the nucleotide uracil takes the place that thymine occupies in DNA. The single strand of mRNA leaves the nucleus through nuclear pores, and migrates into the cytoplasm.

The first product of transcription differs in prokaryotic cells from that of eukaryotic cells, as in prokaryotic cells the product is mRNA, which needs no post-transcriptional modification, whereas, in eukaryotic cells, the first product is called primary transcript, that needs post-transcriptional modification (capping with 7-methyl-guanosine, tailing with a poly A tail) to give hnRNA (heterophil nuclear RNA). hnRNA then undergoes splicing of introns (noncoding parts of the gene) via spliceosomes to produce the final mRNA.

TRANSLATION

Diagram showing the process of translation

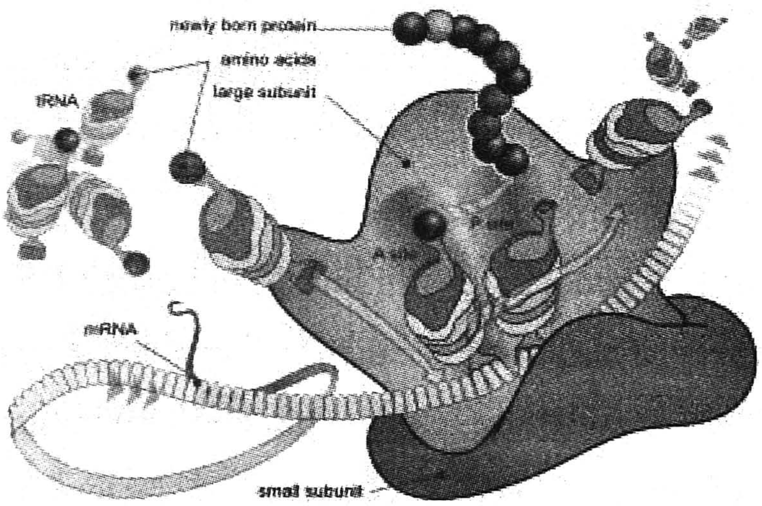


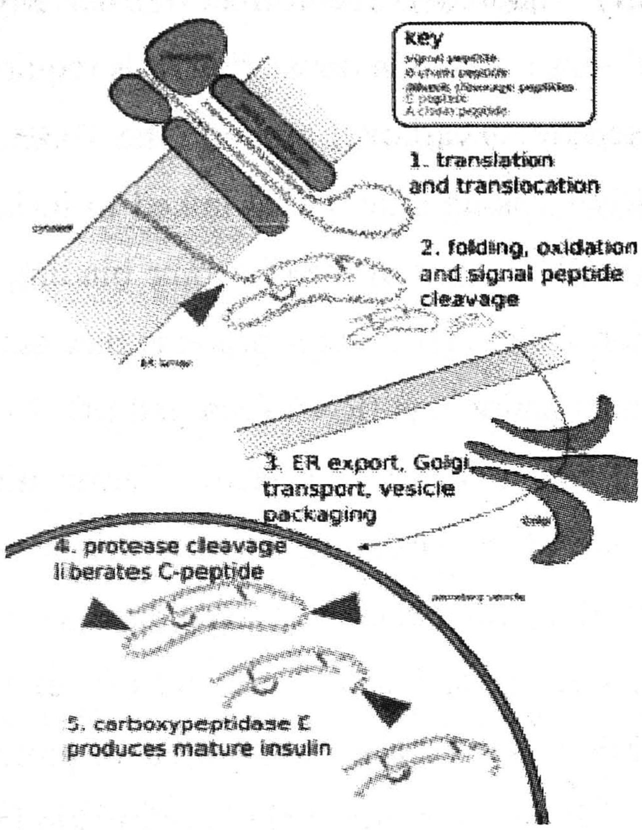
Diagram showing the translation of mRNA and the synthesis of proteins by a ribosome

The synthesis of proteins is known as translation. Translation occurs in the cytoplasm, where the ribosomes are located. Ribosomes are made of a small and large subunit that surround the mRNA. In translation, messenger RNA (mRNA) is decoded to produce a specific polypeptide according to the rules specified by the trinucleotide genetic code. This uses an mRNA sequence as a template to guide the synthesis of a chain of amino acids that form a protein. Translation proceeds in four phases: activation, initiation, elongation, and termination (all describing the growth of the amino acid chain, or polypeptide that is the product of translation).

In activation, the correct amino acid (AA) is joined to the correct transfer RNA (tRNA). While this is not, in the technical sense, a step in translation, it is required for translation to proceed. The AA is joined by its carboxyl group to the 3' OH of the tRNA by an ester bond. When the tRNA has an amino acid linked to it, it is termed "charged". Initiation involves the small subunit of the ribosome binding to 5' end of mRNA with the help of initiation factors (IF), other proteins that assist the process. Elongation occurs when the next aminoacyl-tRNA (charged tRNA) in line binds to the ribosome along with GTP and an elongation factor. Termination of the polypeptide happens when the A site of the ribosome faces a stop codon (UAA, UAG, or UGA). When this happens, no tRNA can recognize it, but releasing factor can recognize nonsense codons and causes the release of the polypeptide chain. The capacity of disabling or inhibiting translation in protein biosynthesis is used by some antibiotics such as anisomycin, cycloheximide, chloramphenicol, tetracycline, streptomycin, erythromycin, puromycin, etc.

POST TRANSLATIONAL MODIFICATIONS

Posttranslational modification (PTM) is a step in protein biosynthesis. Proteins are created by ribosomes translating mRNA into polypeptide chains. These polypeptide chains undergo PTM, (such as folding, cutting and other processes), before becoming the mature protein product.



Post-translational modification of insulin. At the top, the ribosome translates a mRNA sequence into a protein, insulin, and passes the protein through the endoplasmic reticulum, where it is cut, folded and held in shape by disulfide (-S-S-) bonds. Then the protein passes through the golgi apparatus, where it is packaged into a vesicle. In the vesicle, more parts are cut off, and it turns into mature insulin.

A protein (also called a polypeptide) is a chain of amino acids. During protein synthesis, 20 different amino acids can be incorporated to become a protein. After translation, the posttranslational modizfication of amino acids extends the range of functions of the protein by attaching it to other biochemical functional groups (such as acetate, phosphate, various lipids and carbohydrates), changing the chemical nature of an amino acid (e.g. citrullination), or making structural changes (e.g. formation of disulfide bridges).

Also, enzymes may remove amino acids from the amino end of the protein, or cut the peptide chain in the middle. For instance, the peptide hormone insulin is cut twice after disulfide bonds are formed, and a propeptide is removed from the middle of the chain; the resulting protein consists of two polypeptide chains connected by disulfide bonds. Also, most nascent polypeptides start with

the amino acid methionine because the "start" codon on mRNA also codes for this amino acid. This amino acid is usually taken off during post-translational modification.

Other modifications, like phosphorylation, are part of common mechanisms for controlling the behavior of a protein, for instance activating or inactivating an enzyme. Post-translational modification of proteins is detected by mass spectrometry or Eastern blotting.

UNIT – IV PLANT BREEDING

Introduction and Scope

Definition:

Plant breeding is the art and science of improving the heredity of plants for the benefit of mankind.

or

Plant breeding deals with the genetic improvement of crop plants also known as science of crop improvement.

or

Science of changing and improving the heredity of plants

or

Plant breeding can be defined as an art, a science, and technology of improving the genetic makeup of plants in relation to their economic use for the mankind. The mail aim of plant breeding aims to improve the characteristics of plants so that they become more desirable agronomically and economically. The specific objectives may vary greatly depending on the crop under consideration.

Objectives of Plant Breeding :

1. **Higher yield:** The ultimate aim of plant breeding is to improve the yield of economic produce. It may be grain yield, fodder yield, fibre yield, tuber yield, cane yield or oil yield depending upon the crop species. Improvement in yield can be achieved either by evolving high yielding varieties or hybrids.

2. **Improved quality**: Quality of produce is another important objective in plant breeding. The quality characters vary from crop to crop. Eg. grain size, colour, milling and backing quality in wheat. Cooking quality in rice, malting quality in barley, size, colour and size of fruits, nutritive and keeping quality in vegetables, protein content in pulses, oil content in oilseeds, fibre length, strength and fineness in cotton.

3. **Abiotic resistance :** Crop plants also suffer from abiotic factors such as drought, soil salinity, extreme temperatures, heat, wind, cold and frost, breeder has to develop resistant varieties for such environmental conditions.

4. Biotic resistance : Crop plants are attacked by various diseases and insects, resulting in considerable yield losses. Genetic resistance is the cheapest and the best method of minimizing such losses. Resistant varieties are developed through the use of resistant donor parents available in the gene pool.

5. Change in maturity Duration / Earliness : Earliness is the most desirable character which has several advantages. It requires less crop management period, less insecticidal sprays, permits new crop rotations and often extends the crop area. Development of wheat varieties suitable for late planting has permitted rice-wheat rotation. Thus breeding for early maturing crop varieties, or varieties suitable for different dates of planting may be an important objective. Maturity has been reduced from 270 days to 170 days in cotton, from 270 days to 120 days in pigeonpea, from 360 days to 270 days in sugarcane.

6. Determinate Growth : Development of varieties with determinate growth is desirable in crops like Mung, Pigeon Pea (*Cajanus cajan*), Cotton (*Gossypium sp.*), etc.

7. Dormancy : In some crops, seeds germinate even before harvesting in the standing crop if there are rains at the time of maturity, e.g., Greengram, Blackgram, Barley and Pea, etc. A period of dormancy has to be introduced in these crops to check loss due to germinatio n. In some other cases, however, it may be desirable to remove dormancy.

8. Desirable Agronomic Characteristics: It includes plant height, branching, tillering capacity, growth habit, erect or trailing habit etc., is often desirable. For example, dwarf ness in cereals is generally associated with lodging resistance and better fertilizer response. Tallness, high tillering and profuse branching are desirable characters in fodder crops.

9. Elimination of Toxic Substances : It is essential to develop varieties free from toxic compounds in some crops to make them safe for human consumption. For example, removal of neurotoxin in Khesari (*Lathyruys sativus*) which leads to paralysis of lower limbs, erucic acid from *Brassica* which is harmful for human health, and gossypol from the seed of cotton is necessary to make them fit for human consumption. Removal of such toxic substances would increase the

nutritional value of these crops.

10.Non-shattering characteristics: The shattering of pods is serious problem in green gram. Hence resistance to shattering is an important objective in green gram.

11.Synchronous Maturity : It refers to maturity of a crop species at one time. The character is highly desirable in crops like Greengram, Cowpea, and Cotton where several pickings are required for crop harvest.

12.Photo and Thermo insensitivity: Development of varieties insensitive to light and temperature helps in crossing the cultivation boundaries of crop plants. Photo and thermo-insensitive varieties of wheat and rice has permitted their cultivation in new areas. Rice is now cultivated in Punjab, while wheat is a major *rabi* crop in West Bengal.

13.Wider adaptability : Adaptability refers to suitability of a variety for general cultivation over a wide range of environmental conditions. Adaptability is an important objective in plant breeding because it helps in stabilizing the crop production over regions and seasons.

14.Varieties for New Seasons : Traditionally Maize is a *kharif* crop. But scientists are now able to grow Maize as *rabi* and *zaid* crops. Similarly, mung is grown as a summer crop in addition to the main *kharif* crop.

Scope of Plant Breeding:

Since the cultivable land is shrinking and there is no scope for increasing the area under cultivation, the only solution to meet the food requirement is by increasing the crop yield through genetic improvement of crop plants. There are two ways by which yield improvement is possible.

1. Enhancing the productivity of crops

This can be done

- by the proper management of soil and crops involving suitable agronomic practices and harvesting physical resources.
- by using high potential crop varieties created by appropriate genetic manipulation of crop plants.

2. Stabilizing the productivity achieved

This is done by using crop varieties that are bred especially for wide adaptation or for specific crop zones to offset the ill effects of unfavourable environmental conditions prevailing in the areas.

Plant breeding, the past, present and future scopes

Indian agriculture remained stagnant particularly during early sixties. Long spells of severe drought and serious out break of disease in some parts of the country led some futurologists to state that a possible doom in India by the end of the decade. However we achieved a break through in crops such as Rice, wheat, Pearl millet, Jowar and Maize.

The *indica* x *japonica* cross derivative ADT 27 (GEB 24 x Norin 10) is the first high yielding rice of Tamil Nadu. The identification of **Dee Gee Woo Gen** and release of **Wonder rice IR 8** (peta x DGWG) changed the scenario from poverty to problem of plenty.

Like wise identification of dwarfing gene in Japanese wheat variety Norin – 10 by Borlaug and breeding of Mexican dwarf wheat varieties led to the release of wheat varieties life Kalyan sona in India.

In peal millet, breeding by male sterile line Tift 23A at Tifton, Georgia by Burton and his coworker and later on its introduction to India led the release of hybrid bajra HB1 to HB4 which increased bajra production many fold.

In Jowar, breeding of first male sterile line combined kafir 60A and its introduction in to India led to the release of first hybrid sorghum CSH 1 (CK 60A x IS 84) during 1970s.

At present we are in search of alternate source of cytoplasm in almost all crops to breed hybrids with new source of cytoplasm to prevent the possibility of appearance of new pest and diseases. Thus the future of plant breeding is a challenging task. The deployment of innovative breeding techniques will be a new tool to assist the conventional breeding techniques.

Undesirable effects

Plant breeding has several useful applications in the improvement of crop plants. However, it has five main undesirable effects on crop plants.

1. Reduction in Diversity : Modern improved varieties are more uniform than land races. Thus plant breeding leads to reduction in diversity. The uniform varieties are more prone to the new races of pathogen than land races which have high genetic diversity.

2. Narrow genetic base : Uniform varieties have narrow genetic base. Such varieties generally have poor adaptability.

3. Danger of Uniformity : Most of the improved varieties have some common parents in the pedigree which may cause danger of uniformity.

4. Undesirable combinations : Sometimes, plant breeding leads to undesirable combinations. The examples of man made crops having undesirable combination of characters are *Raphanobrassica* and Pomato.

5. Increased susceptibility to minor diseases and pests : Due to emphasis on breeding for resistance to major diseases and insect pests often resulted in an increased susceptibility to minor diseases and pests. These have gained importance and, in some cases, produced severe epidemics. The epidemic caused by *Botrytis cinerea* (grey mold) in chickpea during 1980-82 Punjab, Haryana. The severe infection by Karnal bunt (*Tilletia sp.*) on some wheat varieties, infestation of mealy bugs in Bt cotton.

METHODS OF PLANT BREEDING

Various approaches (*viz.,* selection, hybridization, mutation, etc) that are used for genetic improvement of crop plants are referred to as plant breeding methods or plant breeding procedures or plant breeding techniques. The choice of breeding methods mainly depends on the mode of pollination, mode of reproduction, gene action and breeding objective of crop species. Plant breeding methods are generally classified on the basis of application of crop improvement (general methods, special methods and population improvement approaches) and hybridization (methods involving hybridization and methods not involving hybridization).

Various breeding procedures that are more commonly used for the genetic improvement of various crop plants are known as general breeding methods. Such breeding methods include introduction, selection (pure line selection, mass

selection, progeny selection), hybridization (pedigree, bulk and backcross methods), heterosis breeding, synthetic and composite breeding. On the other hand, those breeding procedures that are rarely used for improvement of crop plants are referred to as special breeding methods. Such methods include: mutation breeding, polyploidy breeding, wide crossing or distant hybridization and biotechnology. Four breeding approaches, *viz.*, recurrent selection, disruptive matin g and selection, diallel selective mating system and biparental mating are used mainly for population improvement.

Classification of Plant Breeding Methods

Basis of classification and

Types of methods

Breeding methods included

A. Application in crop improvement

a. General Methods Plant introduction, Pure line selection, mass selection, progeny selection, pedigree method, bulk method, back cross method, SSD, clonal selection, heterosis breeding, synthetics and composites.
b. Special Methods Mutation breeding, Polyploidy breeding, transgenic breeding, molecular breeding.
c. Population Improvement Recurrent selection, disruptive selection, diallel selective approaches mating system, biparental mating.

B. Hybridization

a. Methods involving hybridization Pedigree, bulk, backcross and SSD Methods: heterosis breeding, and population improvement approaches and molecular breeding (marker aided selection).
b. Methods not involving hybridization Plant Introduction, pureline selection, mass selection, progeny selection, clonal selection, mutation breeding and transgenic breeding. There are some differences in the breeding methods used for self pollinated and cross pollinated species. Self pollinated species are homozygous, hence we can start hybridization directly. Cross pollinated species, on the other hand, are highly heterozygous. Hence we can not start

hybridization directly. First we have to develop inbred lines by selfing or inbreeding and then only hybridization can be taken up. We have to exploit homozygosity in self pollinated crops and heterozygosity in cross pollinated species. Asexually propagated species such as sugarcane, potato, sweet potato, etc., are highly heterozygous. Hence, F1 hybrids in such crops exhibit segregation and selection can be practiced in *F1* generation. The superior clones are identified and further multiplied. The maintenance or conservation of hybrid vigour is easy in such crops because of asexually propagation.

Methods of Breeding Autogamous species

Plant breeding methods that are used for genetic improvement of self pollinated or autogamous species include:

1. Plant Introduction
2. Pureline selection
3. Mass selection
4. Pedigree method
5. Bulk method
6. Single seed descent method
7. Backcross method
8. Heterosis breeding
9. Mutation breeding
10. Polyploidy breeding
11. Distant hybridization
12. Transgenic breeding.

Four breeding approaches, *viz.,* recurrent selection, disruptive selection, diallele selective mating, and biparental mating are used for population improvement.

Methods or Breeding Allogamous species

Breeding methods that are used for genetic improvement of cross pollinated or allogamous species include

(1) Plant introduction

(2) Mass and progeny selection

(3) Backcross method

(4) Heterosis breeding

(5) Synthetic breeding

(6) Composite breeding

(7) Polyploidy breeding

(8) Distant hybridization

(9) Transgenic breeding

Mutation breeding is rarely used in allogamous species. Three breeding approaches *viz.,* recurrent selection, disruptive mating and biparental meting are used for population improvement.

Methods of Breeding Asexually Propagated Species

Important breeding methods applicable to asexually propagated species are

(1) Plant Introduction

(2) Clonal selection

(3) Mass selection

(4) Heterosis breeding

(5) Mutation breeding

(6) Polyploidy breeding

(7) Distant hybridization

(8) Transgenic breeding.

Mass selection is rarely used in asexually propagated species.

Brief account of breeding methods

Plant introduction is applicable to all three groups of crop plants, *viz.,* self pollinated, cross pollinated and asexually propagated species. It is an old est and rapid method of crop improvement. The introduced material may be used in three ways *viz.,*

(1) Directly as a variety

(2) As a variety after selection

(3) As a parent in the hybridization for development of variety or hybrid Pureline selection is applicable to self pollinated species. It is also used sometimes in cross pollinated species for development of inbred lines. A single best pure line is

released as a variety. Thus a pureline variety is homozygous and homogeneous population.

Mass selection is common in cross pollinated species and rare in self pollinated and asexually propagates species. In self pollinated crops, a mass selected variety is a mixture of several purelines. Thus it is a homozygous but heterogeneous population. In cross pollinated species, a mass selected variety is a mixture of several hetero and homozygotes. Thus, it is a heterozygous and heterogeneous population.

Progeny selection is used in cross pollinated species. A variety developed by this method is heterozygous and heterogeneous population because it consists of several hetero and homozygotes.

Pedigree method is applicable to both self and cross pollinated species. In self pollinated crops progeny of a single best homozygote is released as a variety. Thus a variety developed by this method has a homozygous and homogeneous population. In cross pollinated species, it is used for developed of inbred lines. Bulk nad single seed descent methods are used in self pollinated species. Progeny of a single best homozygote is released as a varie ty by these methods. Thus, varieties developed by these methods are homozygous and homogeneous.

Backcross method is applicable in all three groups of crop species. This method is used for transfer of oligogenic characters from a donor source to a well adapted variety. This method is also used for development of multilines, Isogenic lines and transfer of male sterility. This method is more effective in transferring oligogenic characters than polygenic traits. The end product of backcross method is similar to parent variety expect for the character which has to be transferred from the donor source.

Multiline varieties are developed in self pollinated species. They are mixture of several Isogenic lines, closely related lines or unrelated lines. Thus a multiline variety is a homozygous but heterogeneous population.

Clonal selection is used in asexually propagated species. In this method progeny of a single best clone is released as a variety. Such variety has heterozygous but homogeneous population.

Heterosis breeding is used in/all the three groups. However, it is common in cross pollinated and asexually propagated species and rare in self pollinated species. A hybrid variety has homogeneous but heterozygous population. Synthetic and composite varieties are developed in cross pollinated species. Such varieties consist of several homozygotes and heterozygotes and thus constitute a heterogeneous population.

Mutation breeding is common in self pollinated and asexually propagated species and rare in cross pollinated species. A mutant variety differs from parent variety in one or few characters. A mutant differs from a segregant in two main ways. Firstly, the frequency of segregants is very high and that of mutant is extremely low (0.1%). Secondly, mutant differs from parent variety in one or few characters, where as a segregant differs from parent material in several characters.

Polyploidy breeding is common in asexually propagated species and rare in self and cross pollinated species. A polyploidy variety differs from parent variety in chromosome numbers and exhibit gigant morphological characters.

Distant hybridization is used in all the three types of crop species. However, this method is used for transferring some desirable genes from wild species to the cultivated ones. Generally, backcross method is used for transfer of oligogenic characters and pedigree method for transfer of polygenic characters.

Transgenic breeding is applicable to all three types of crop species. This method is used to solve specific problems which can not be solved by conventional breeding techniques. This method will serve as a tool and can not be used as a substitute for conventional breeding methods.

Recurrent selection is common in cross pollinated species and rare in other two groups. It is used for accumulating favourable genes in a population *i.e.,* for population improvement. Other approaches which are used for population improvement include disruptive mating, diallel selective mating (DSM) and biparental mating. DSM is used in self pollinated species and other two techniques can be used both in self and cross pollinated species.

PURE LINE THEORY

A pure line is the progeny of a single self fertilized homozygous plant. The concept of pureline was proposed by **Johannsen** on the basis of his studies with beans *(Phaseolus vulgaris)* variety called Princess. He obtained the seeds from the market and observed that the lot consisted of a mixture of larger as well as smaller size seeds. Thus there was variation in seed size. Johannsen selected seeds of different sizes and grown them individually. Progenies of larger seeds produced larger seeds and progenies from smaller seeds produced small seeds only.

This clearly showed that there is variation in seed size in the commercial lot and it has a genetic basis. He studied nineteen lines al together. He concluded that the market lot of the beans is a mixture of pure lines. He also concluded whatever variation observed with in a line is due to environment only. Confirmatory evidence was obtained in three ways. In line 13 which is having 450 mg seed wt he divided the seeds on weight basis. He divided the line into seeds having 200, 300, 400 and 500 mg weights and studied the progenies. Ultimately he got lines having weight ranging from 458 to 475. Thus the variation observed is purely due to environment.

The second evidence was that selection with in a pure line is ineffective. From a pure line having 840 mg selection was made for large as well as small seeds. After six generations of selection the line for large seed as well as for small seed gave progenies having 680-690 mg. Thus it was proved that selection within a pure line is ineffective. In third evidence when parent - offspring regression was worked in line thirteen. It worked to zero indicating that variation observed is non heritable and it is due to environment only.

ORIGIN OF VARIATION IN PURELINES

1. Mechanical mixtures.
2. Natural hybridization.
3. Chromosomal aberrations.
4. Natural mutation.
5. Environmental factors.

EFFECT OF SELF-POLLINATION ON GENOTYPE

Self-pollination increases homozygosity with a corresponding decrease in heterozygosity. For example an individual heterozygous for a single gene Aa is self pollinated in successive generations, every generation of selfing will reduce the frequency of heterozygote Aa to 50 percent of that in the previous generation. There is a corresponding increase in homozygotes AA and aa. As a result, after 10 generations of selfing virtually all the plant in the population will be homozygous AA and aa.

No. of generations of selfing	Frequency(%) AA	Aa	aa	Frequency (%) Homozygote	Heterozygote
0	0	100	0	0	100
1	25	50	25	50	50
2	25 + 12.5	25	25 + 12.5	75	25

This can be calculated by the formulae = $[2^m - 1) / 2^m]^n$

where m = No. of generations of self-pollination.

n = No. of genes segregating.

When number of genes are segregating together, each gene would become homozygous at the same rate as Aa. Thus the number of genes segregating does not affect the percentage of homozygosity. Similarly linkage between genes does not affect the percentage of homozygosity in the population.

Genetic advance under selection

Normally selection is practiced based on the phenotype of the individual plant. The phenotype in turn is the result of joint action of genotype and environment i.e.,

$V_P = V_G + V_E$

WHERE P= PHENOTYPE

G = GENOTYPE

E = Environment

Genetic advance:

It is the improvement in the mean genotypic value of the selected families over the base population is known as genetic advance under selection. Genetic advance under selection depends upon,

- Genetic variability among different plants or families in the base population.
- The herbitability of the character under selection.
- The intensity of selection i.e., the proportion of plants or families selected.

Genetic advance under selection may calculated as follows.

$$GS = (K)\ (\sigma P)\ (H)$$

Where GS = Genetic advance under selection

k = Selection differential

σP = Phenotypic standard deviation of the base population

H = Heritability of the character under selection.

PURELINE SELECTION

A large number of plants are selected from a self pollinated crop. The selected plants are harvested individually. The selected individual plants are grown in individual rows and evaluated and best progeny is selected, yield tested and released as a variety.

CHARACTERISTICS OF PURELINES

1. All plants within a pure line have the same genotype.
2. The variation with in a pureline is environmental and nonheritable.
3. Purelines become genetically variable with time due to natural hybridization, mutation and mechanical mixtures.

General steps for making a pureline selection

First Season : From the base population select best looking plants having the desirable characters. Harvest them on single plant basis.

Second Season : The selected single plants are grown in progeny rows and estimate the performance. Reject unwanted progenies.

Third Season : Repeat the process of second season.

Fourth Season : Grow the selected single plants in replicated preliminary yield trial along with suitable check or control variety.

Fifth Season : Conduct regular comparative yield trial along with check variety and select the best culture.

Sixth Season : Conduct multilocation trial in different research stations along with local check.

Seventh Season : Conduct Adaptive Research Trial in farmer's field. Fix the best yielder and release it as a variety thro' Variety Release committee.

Advantage of Pureline Selection.

1. The purelines are extremely uniform since all the plants in the variety will have the same genotype.
2. Attractive and liked by the farmers and consumers.
3. Purelines are stable and long last for many years.
4. Due to its extreme uniformity the variety can be easily identified in seed certification programmes.

Disadvantages :

1. It does not have wide adaptability because improvement is made only in the local variety.
2. Time required for developing a variety is more when compared to mass selection.
3. Depending on the genetic variability present in the base population only the improvement is made. If there is no genetic variability improvement cannot be made.
4. Breeder has to spend more time compared to mass selection.

Achievements :

Several varieties developed by pureline selection were released in many crops. Some examples are given below

Rice : Mtu-1, Mtu-3

Sorghum : G 1, G 2, M 1

Groundnut : TMV 3, TMV 4, Kadiri 71-1

Redgram : TM-1, ST-1

Chillies : G1 & G2

Ragi : AKP 1, AKP 7

MASS SELECTION

It is the earliest method of selection. Man has always practiced mass selection consciously or unconciously from the time of domestication. In its most basic form mass selection consists of selecting individuals on the basis of phenotypic superiority and mixing the seeds for using as planting material for next season.

Procedure for evolving variety by mass selection

First year : Large number of phenotypic ally similar plants having desirable characters are selected. The number may vary from few hundred to few thousand. The seeds from the selected plants are composited to rise the next generation.

Second year: composited seed planted in a preliminary field trial along with standard checks. The variety from which the selection was made should also be included as check. Phenotypic characteristics of the variety are critically examined and evaluated.

Third to sixth year: The variety is evaluated in coordinated yield trials at several locations. It is evaluated in an initial evaluation (IET) trial for one year. If found superior it is promoted to main yield trials for 2 or 3 years.

Seventh year : if the variety is proved superior in main yield trials it is multiplied and released after giving a suitable name.

Modification of mass selection

Mass selection is used for improving a local variety. Large number of plants are selected (I year) and individual plant progenies are raised (II year). Inferior, segregating progenies are reflected. Uniform, superior rows are selected and the seed is bulked. Preliminary yield trials are conducted in third year. Fourth to seventh year multi location tests are conducted and seed is multiplied in eight year and distributed in ninth year. Many other modifications also are followed depending on the availability of time and purpose for which it is used.

Merits of Mass Selection:

a. Can be practiced both in self and cross pollinated crops
b. The varieties developed through mass selection are more widely adopted than pure lines.
c. It retains considerable variability and hence further improvement is possible in future by selection
d. Helps in preservation of land races
e. Useful for purification of pureline varieties
f. Improvement of characters governed by few genes with high heritability is possible.
g. Less time consuming and less expensive.

Demerits of mass selection

a. Varieties are not uniform
b. Since no progeny test is done, the genotype of the selected plant is not known
c. Since selection is based on phenotype and no control over pollination the improvement brought about is not permanent. Hence, the process of mass selection has to be repeated not and then.
d. Characters which are governed by large number of genes with low heritability can not be improved.
e. It can not create any new genotype but utilizes existing genetic variability.

Achievements

Mass selection must have been used by pre-historic man to develop present day cultivated cross from their wild parents. It was also used extensively before pureline selection came into existence.

Cotton : Dharwad American Cotton

Groundnut : TMV-1 & TMV-2

Bajra : Pusa Moti, Baja Puri, Jamnagar Giant, AF3

Sorghum : R.S. 1

Rice: SLO 13, MTU-15

Potato: K122

COMPARISON BETWEEN PURELINE AND MASS SELECTIONS

	Pureline selection	Mass selection
1.	The new variety is a pureline	The new variety is a mixture of purelines.
2.	The new variety is highly uniform. In fact, the variation within a pureline variety is purely environmental.	The variety has genetic variation of quantitative characters, although it is relatively uniform in general appearance.
3.	The selected plants are subjected to progeny test.	Progeny test is generally not carried out.
4.	The variety is generally the best pureline present in the original population. The pureline selection brings about the greatest improvement over the original variety.	The variety is inferior to the best pureline because most of the purelines included in it will be inferior to the best pureline.
5.	Generally, a pureline variety is expected to have narrower adaptation and lower stability in performance than a mixture of purelines.	Usually the variety has a wider adaptation and greater stability than a pureline variety.
6.	The plants are selected for the desirability. It is not necessary they should have a similar phenotype.	The selected plants have to be similar in phenotype since their seeds are mixed to make up the new variety.
7.	It is more demanding because careful progeny tests and yield trials have to be conducted.	If a large number of plants are selected, expensive yield trials are not necessary. Thus it is less demanding on the breeder.

Pedigree Method

In the pedigree method, individual plants are selected from F2 and subsequent generations, and their progenies are tested. During the entire operation a record of all parent off spring relationships is kept. This is known as pedigree record. Individual plant selection is continued till the progenies show no segregation. At this stage the selection is done among the progenies, multilocation tests are conducted and released as varieties.

The pedigree may be defined as a description of the ancestors of an individual and it generally goes back to some distant ancestors. It is useful to know the relationship of two individuals and useful for selection of parents and prediction of outcome of the cross. Procedure of pedigree method

1st year : cross is made between the parents possessing desirable characters.

2nd year : Sow the F1 seed giving wide spacing so that each F1 plant produces more seeds. Raise as many F1 plants as possible to produce large number of F2 seeds. Harvest in bulk.

3rd year : Grow 2000-10000 plants of F2 giving wide spacing for full expression of the characters in F2 generation plants. Grow parents for comparision. Depending upon the facilities and objectives of the programme about 100-500 superior plants are selected. The value of selection depend on the skill of the breeder. He has to judge which F2 plant will produce superior progeny for characters under consideration. The breeder develops this skill through close study of the crop for many generations. The selection in F2 is done for simply inherited characters like head type disease resistance etc. and selection for characters governed by many genes like yield will be reserved for later generations. The selected plants are harvested separately and given serial numbers and description entered in pedigree registers.

4th year : Progeny rows of F3 i.e. seeds of one selection plant in one row are space planted along with parents and checks. From superior progeny rows, individual plants with desirable characters are selected (about 50-100 families and about 5 plants in each family and harvested separately). Diseased, lodging and undersirable progenies are discarded.

5th year : F4 plants raised again as head to row. Desirable plants are selected from desirable
rows and harvested separately.

6th year : F5 plants raised in 3 row plots i.e. seeds of each selected plant sown in 3 rows. By this time many families might have become reasonably homozygous. For comparison check variety is grown for every 3 or 5 block. Progenies are evaluated for yield and the inferior ones are rejected. The number should be reduced to 25-50. superior plants from superior progenies are selected. Plants from each progeny are bulked.

7th year : F6 individual plant progenies are grown in multi-row plots and evaluated. Inferior progenies are rejected and superior progenies are selected. Plants of each progeny are harvested in bulk. Diseased and inferior plants from the progenies are removed.

8th year : F7 preliminary yield trial with 3 or more replications ar conducted to identify superior lines. The progenies are evaluated for many characters including yield. Standard commercial varieties must be included as checks. Two to five outstanding lines are selected and advanced to coordinated yield trials.

9th, 10 th & 11th year : selected lines are tested in several localities for 2 or 3 years for adaptation tests. Lines are evaluated for all characters mainly yield and disease resistance. A line that is superior to commercial variety in yield and other characters is selected.

12th year : Selected superior lines is named, multiplied and released as a new variety. Number of year can be reduced if generations are advanced during off seasons either in green house or under irrigated conditions. Several modifications for the above described pedigree method are followed by breeders depending upon the crop, time and availability of funds and facilities like labour, land etc.

Early generation tests :

The objective of these test is to find out superior crosses and superior progenies in early generations i.e. in F2 and F3. we need not advance all the crossed and all selected progenies in each cross upto F8. much labour, time and cost would be saved by this early generation testing. A more reliable information about the

potential crosses and progenies may be obtained by conducting replicated tests (preferably in more location) and evaluating them for yield and other characters in F2 or F3 itself. A desirable cross or progeny should have high mean yield, high genetic variance and high expected genetic advance under selection. Other crosses and progenies are rejected in the beginning i.e. F2 and F3 generations itself.

F2 progeny testing : Another modification for pedigree method. In F2 make as many single plants selections as possible. From F3 to F6 advance the progenies in bulk making selections of the progenies as a wholw and discarding the inferior progenies. Thus each of the progeny is derived from the single plant selected in F2 generation. In F6 make single plant selections in each of the progeny. Compare the yields of the single plants with progenies from which they are selected. Select superior single plant progenies and advance to preliminary yield trials, multilocation tests etc. There are two advantages 1. No. of crosses can be handled simultaneously 2. Natural selection operates from F 3 to F6 since they are advanced in bulk.

Mass pedigree method : This is another modified pedigree method. Crosses are made and further generations grown in bulk or as mass until suitable season occurs for making desirable selections against drought, insect and diseases etc. The population will be exposed to the natural conditions of vagaries. From the remaining population individual plants are selected and harvested progenies are evaluated for yield and other characters in preliminary yield trials and further generations are proceeded as in pedigree method till release of variety. The advantages of both bulk and pedigree methods can be obtained and large number of crosses can be handled at a time. The disadvantage is that it takes a bit longer time.

Merits of pedigree method :

a. It gives maximum opportunity for the breeder to use his skill and judgment in selection of plants
b. It is well suited for the improvement of characters which can be easily identified and are simply inherited.

c. Transgressive segregation for yield and other quantitative characters may be recovered.

d. Information about the inheritance of characters and pe digree of lines can be obtained.

e. Inferior plants and progenies are eliminated in early generations.

f. It takes less time than bulk method to develop new variety.

Demerits of pedigree method :

a. Valuable genotypes may be lost in early generations, if sufficient skill and knowledge are lacking in the breeder, at the time of selection.

b. No opportunity for natural selection

c. Difficult to handle many crosses

d. Maintenance of records, selections, growing progeny rows etc are time consuming and laborious.

Achievements: Large number of varieties have been developed by pedigree method in many crops. A few examples are

Wheat – NP-52

Rice – ADT – 25, Jaya, Padma

Cotton – Lakshmi, Digvijay,

Sorghum – Co 18, Rs 610

Tobacco – NP 222

BULK METHOD

The bulk method was first proposed by Nilsson Ehle in 1908 at Svalof. This method is also known as **mass method** 'or' Population method of breeding. Steps involved in bulk breeding method are,

- Isolation of Homozygus lines
- Waiting for the opportunity for selection
- Opportunity for natural selection.
- F2 and subsequent generations are harvested in mass as bulk to raise the next generation.

- At the end of the bulking period (after attaining homozygosity) individual plants are selected and evaluated similar manner as pedigree method of breeding.

THE PROCEDURE FOR BULK METHOD

The exact procedure for the bulk method would vary depending upon the objective of breeder. The following procedure is described for the isolation of homozygous lines. The breeder may introduce various modifications in the scheme to suit his needs.

Hybridization : Parents are selected according to the objective of the breeding programme. A simple or a complex cross is then made depending upon the number of parents involved.

F1 Generation : F1 is space-planted and harvested in bulk. The number of F1 plants should be as large as possible ; usually more than 20 plants should be grown.

F2-F6 Generations : F2 to F6 generations are planted at commercial seed rates and spacings. These generations are harvested in bulk. During this period, environmental factors, disease and pest outbreaks would change, the frequencies of different genotypes in the population. Artificial selection is generally not done. The population size should be as large as possible, preferably 30,000-50,000 plants in each generation.

F7 Generations : About 30-50 thousand plants are space-planted. 1000 to 5000 plants with superior phenotypes are selected and their seeds harvested separately. Selection is based on the phenotypes of plants, grain characteristics, disease reaction, etc.

F8 Generation : Individual plant progenies are grown in single or multi-row plots. Most of the progenies would be reasonably homozygous and are harvested in bulk. Weak and inferior progenies are rejected on the basis of visual evaluation. Only 100-300 plant progenies with desirable characteristics are saved. Some progenies which show segregation are generally rejected unless they are of great promise. In promising pr ogenies, individual plants may be selected ; preliminary yield trial will be delayed for one year in such cases.

F9 Generation : Preliminary yield trial is conducted by using standard commercial varieties as checks. The progenies which are superior than the check are advanced. Quality test may be conducted to further reject undesirable progenies. The progenies are evaluated for height, lodging resistance, maturity date, disease resistance and other important characteristics of the crop species.

F10-F13 Generations : Replicated yield trials are conducted over several locations using standard commercial varieties as checks. The lines are evaluated for important characteristics in addition to yield, disease resistance and quality. If a line is superior to the standard varieties in yield trials, it would be released as a new variety.

F14 Generation: Seed of the released variety is increased for distribution to the cultivators.

MERITS OF BULK METHOD

a. The bulk method is simple, convenient and less expensive.
b. Since, each F2 plant is equally represented till F6, no chance of elimination of good genotypes in early generations.
c. Artificial or natural disease epiphytics, winter killing high temperature etc. eliminates undesirable types and increases the frequency of desirable type. Thus isolation of desirable types becomes easier.
d. Progenies select from long term bulks are superior than the selection from F2 or short term bulk.
e. Since, little work and attention is needed in F2 and subsequent generation more no. of crosses can be handled.
f. No pedigree records which saves time
g. Since large population are grown, transgressive segregants are more likely to appear and increase due to natural selection. Hence, there is a greater chance to isolate good segregants than pedigree method.

DEMERITS OF BULK METHOD

a. The major disadvantage of bulk method is that it takes a much longer time to develop a new variety. Natural selection becomes important only after F8 or F10, and bulking may have to be done upto F20 or more. Thus the time required is considerably longer, and most breeders do not use the bulk method simply for this reason.

b. In short-term bulks, natural selection has little effect on the genetic composition of populations. But short -term bulks are useful for the isolation of homozygous lines and for specific objectives as in Harlan's mass - pedigree method.

c. It provides little opportunity for the breeder to exercise his skill or judgement in selection. But in the modified bulk method, the breeder has ample opportunity for practicing selection in the early segregating generations.

d. A large number of progenies have to be selected at the end of the bulking period.

e. Information on the inheritance of characters cannot be obtained which is often available from the pedigree method.

f. In some cases, at least, natural selection may act against the agronomically desirable types.

Comparison between Pedigree and Bulk Method

S. No.	Pedigree Method	Bulk Method
1	Most widely used Breeding method	Used only to a limited extent
2	Individual plants are selected in F2 and subsequent generations and individual plant progenies are grown	F2 and subsequent generations are grown in bulk
3	Artificial selection ; artificial	Mainly natural selection. In certain

	disease epidemics etc. are an integral part of the method	cases artificial selection may be essential
4	Natural selection does not play any role	Natural selection determines the composition of the pop n at the end of the bulking period
5	Pedigree Records have to be maintained which is often time consuming and laborious	No pedigree records are maintained
6	Generally its taken 12-13 years to release a new variety	Takes more than 15 years.
7	Requires close attention of breeder from F2 onwards	It is quite simple and does not require much attention
8	Planting (spacing) the segregating generations are space planted to permits effective individual plant selection	The bulk populations are generally planted at commercial planting rates
9	Population size is small in comparison to bulk	The population size is large

BACKCROSS METHOD

In backcross method of breeding, the hybrid and the progenies in subsequent generations are repeatedly backcrossed to one of the parents. As a result, the genotype of the backcross progeny becomes increasingly similar to that of the recurrent parent. The objective of backcross method is to improve one or two specific defects of a high yielding variety.

Pre-requisite for back cross breeding

1. A suitable recurrent parent must be available which lacks in one or two characteristics.
2. A suitable donor parent must be available
3. The character to be transferred must have high heritability and preferably it should be determined by one or two genes.

4. A sufficient number of back crosses should be made so that the genotype of recurrent parent is recovered in full.

Application of back cross method

This method is commonly used to transfer disease resistance from one variety to another. But it is also useful for transfer of other characteristics.

a. Intervarietal transfer of simply inherited characters

E.g. Disease resistance, Seed coat colour

b. Intervarietal transfer of quantitative characters.

E.g. Plant height, Seed size, Seed shape.

c. Interspecific transfer of simply inherited characters.

E.g. Transfer of disease resistance from related species to cultivated species.

E.g. Resistance to black arm disease in cotton from wild tetraploid species into *G.hirsutum*

4. Transfer of cytoplasm :

This is employed to transfer male sterility. The female parent will be having the sterile cytoplasm and recurrent parent will be used as male parent.

E.g. *Sesamum malabariucum* x *S.indicum*

Female parent Recurrent parent.

5. Transgressive segregation :

Back cross method may be modified to produce transgressive segregants. Th F_1 is back crossed to recurrent parent for 2 to 3 times for getting transgressive segregants.

6. Production of isogenic lines.

7. Germplasm conversion : E.g. Production of photo insensitive line from photo Sensitive germplasm thro' back crossing. This was done in the case of sorghum. Popularly known as conversion programme.

Procedure for backcross method

The Plan of backcross method would depend upon whether the gene being transferred is recessive or dominant. The plan for transfer of a dominant gene is simpler than that for a recessive gene.

First Year	Non recurrent Parent B			Recurrent Parent A	
	RR		x	rr	
	Resistant to rust			Susceptible to rust	
	F_1	Rr	x	rr	BC_1
		Resistant.			
~~rr~~		Rr	x	rr	BC_2
~~rr~~		Rr	x	rr	BC_3
~~rr~~		Rr	x	rr	BC_4
~~rr~~		Rr	x	rr	BC_5

Back cross up to 6th or 7th generation. After 7th Backcross rust resistant lines are self pollinated. Harvest is done on single plant basis.

8th Season

a) Individual plant progenies grown

b) Homozygous plants having resistance and resembling parent A are selected harvested and bulked.

9th Season - Yield trials.

10th Season - Seed multiplication and distribution.

Steps

First Season

Hybridization : Crossing between parent B donor (Female) and Susceptible parent A recipient (male)

Second Season :

Raising the F_1 and backcrossed to recurrent parent A.

Third season :

Growing the BC_1F_1. It will be segregating for 1 susceptible (rr) : 1 resistant (Rr). Rust resistant plants are backcrossed with recurrent parent A. This is second backcross.

Fourth Season

Raising BC_2 F_1 will again segregate in the ratio of 1:1. Third backcross is done with resistant plants.

Fifth Season to Eighth Season

Raising backcross F_1s and crossing resistant plants with recurrent parent is continued up to 7th $backcross_1$

Ninth season :

Raising BC_7F_1 and observing resistant lines. The plants resembling parent A coupled with resistance is selected and harvested on single plant basis.

Tenth Season :

Growing the progeny rows and observing each row for resistance. Best rows are selected and harvest is done on row basis

Eleventh Season :

The row bulk is raised in yield trial along with check, the best plots are selected.

Twelfth season :

Selected plot seeds are multiplied and released as new variety.

Back Cross Method - Transfer of Recessive Gene

I Season Non recurrent parent B Recurrent parent A Hybridization Resistan

Susceptible rr x RR F_1 Rr

II Season Grow the F_1 Rr (Selfing)

III. Season Grow F_2 RR /Rr : rr x RR BC_1

IV Season Grow BC_1 F_1 Rr (Selfing)

V Season Grow BC_2 F_2 RR : Rr : rr x RR BC_2

VI Season Grow BC_2 F_1 Rr (Selfing)

VII Season Grow BC_2 F_2 /RR : /Rr : rr x RR BC_3

VIII Season Raise BC_3F_1

IX Season Raise BC_3F_2 and it will segregate in to 1:2:1 with resistant segregant make Backcross 4 (BC_4)

X Season Do as on VIII Season

XI Season Do as in IX season.

Continue this process till 7th or 8th backcross. After studying 8th BCF_2 select plants resembling parent B coupled with resistance. Harvest them on single plant basis. Next season raise them in progeny rows and select best progenies. Compare them in yield trial and fix the best culture, multiply it and release it as a variety. While selecting plants artificial bombardment for disease is to be done.

Steps :

I Season : Make a cross between donor parent A and recurrent parent B and Harvest the hybrid. The donor parent A is resistant which is governed by recessive genes. The susceptibility is governed by dominant gene in parent B.

II Season : Grow the F_1 which will be susceptible - Harvest them.

III Season : Grow F_2 which will be segregating in the ratio of 1:2:1 i.e. 3/4 susceptible and 1/4 resistant. With the resistant lines (rr) make first backcross with parent A having dominant RR gene. Harvest BC_1 F_1

IV Season : Grow BC_1F_1

V season : Grow BC_1 F_2 which will be segregating as we saw in III season. Repeat the process of third season. This will give BC_2F_1

VI Season : Grow BC_2 F_1

VII Season : Grow BC_2 F_2 then repeat the process of V season. This will give BC_3 F_1.

VIII Season : Grow BC_3 F_1

IX Season : Grow Bc_3 F_2 and repeat the process of VII season. Harvest BC_4 F_1

X Season : Grow BC_4 F_1

XI Season : Grow BC_4 F_2 and repeat the process of IX Season. Harvest BC_5F_1

XII, XIII : Repeat the process and carry out backcross up to 7 time.

XIV, XV Season : While studying BC_7 F_2

Select single plants having resistance and resembling parent B.

XVI Season : Study the progenies in progeny rows and select best progenies.

XVII Season : Conduct yield trial and select best material.

XVIII Season : Multiply the seeds and distribute it as improved variety with resistance to disease.

Note: While studying Back cross F_2s they should be bombarded with artificial epiphytotic conditions.

MERITS OF BACKCROSS METHOD

1. The genotype of the new variety is nearly identical with that of the recurrent parent, except for the genes transferred. Thus the outcome of a backcross programme is known beforehand, and it can be reproduced any time in the future.
2. It is not necessary to test the variety developed by the back cross method in extensive yield tests because the performance of the recurrent parent is already known. This may save upto 5 years time and a considerable expense.
3. The backcross programme is not dependent upon environment, except for that needed for the selection of the character under transfer. Therefore, off-season nurseries and green - houses can be used to grow 2-3 generation each year. This would drastically reduce the time required for developing the new variety.
4. Much smaller population are needed in the backcross method than in the case of pedigree method.
5. Defects, such as, susceptibility to disease, of a well-adapted variety can be removed without affecting its performance and adaptability. Such a variety is often preferred by the farmers and the industries to an entirely new variety because they know the recurrent variety well.
6. This is the only method for interspecific gene transfers.
7. It may be modified so that transgressive segregation may occur for quantitative characters.

DEMERITS OF BACKCROSS METHOD

1. The new variety generally cannot be superior to the recurrent parent, except for the character that is transferred.
2. Undesirable genes closely linked with the gene being transferred may also be transmitted to the new variety.
3. Hybridization has to be done for each backcross. This is often difficult, time taking and costly.

4. By the time the backcross is over, the recurrent parent may have been replaced by other varieties superior in yielding ability and other characteristics.

CLONAL SELECTION

It is one of the common method of breeding for vegetatively propagated crops

Clone : A clone is a group of plants produced from a single plant through asexual reproduction.

The crop plants can either be propogated by seeds or by vegetative parts. The vegetative propogation is resorted due to :

a. Lack of seed : Eg. Ginger, termiric
b. There is short viability of seed : Eg. Sugarcane
c. The seed production is very rare : Eg. Banana
d. Seeds are produced under special conditions only : Eg. Sugarcane, potato

Characteristics of Asexually propagated crops :

a. Majority of them are perennials : Eg . Sugarcane, fruit trees.
b. The annual crops are mostly tuber crops : Eg. Potato, cassava, Sweet potato
c. Many of them show reduced flowering and seed set
d. They are invariably cross pollinated
e. These crops are highly heterozygous and show severe inbreeding depression upon selfing.
f. Majority of asexually propagated crops are polyploids : Eg. Sugarcane, Potato, Sweet,
g. Potato
h. Many species are interspecific hybrids. Eg. Banana, Sugarcane

Characteristics of a clones :

a. All the individual belonging to a single clone are identical in genetype
b. The phenotypic variation within a clone in due to environment only
c. The phenotype of a clone is due to the effects of genotype(g), the environment(e) and the genotype x environment interaction (GxE), over the population mean(M)

d. Theoratically clones are immortal. They deteriorate due to viral/bacterial infection and mutations.
e. Clones are highly heterozygous and stable
f. They can be propagated generation after generation without any change.

Importance of a clone

a. Owing to heterozygosity and sterility in many crops clones are the only means of propagation.
b. Clones are used to produce new varieties.
c. Clones are very useful tools to preserve the heterozygosity once obtained. In many crops the superior plants are maintained. (Mango, orange, apple, sugarcane)

Sources of clonal selection :

1. Local varieties
2. Introduced material
3. Hybrids and
4. Segregating populations

The various steps involved in clonal selection are briefly mentioned below.

First year : From a mixed variable population, few hundred to few thousand desirable plants are selected. Rigid selection can be done for simply inherited characters with high heritability. Plants with obvious weakness are eliminated.

Second year : Clones from the selected plants are grown separately, generally without replication. This is because of the limited supply of propagating material for each clone, and because of the large number of the clones involved. Characteristics of the clones will be more clear now than in the previous generation. Based on the observations the inferior clones are eliminated. The selection is based on visual observations and on judgement of the breeder on the value of clones. Fifty to one hundred clones are selected on the basis of clonal characteristics.

Third year : Replicated preliminary yield trial is conducted. A suitable check is included for comparison few superior performing clones with desirable characteristics are selected for multilocation trials. At this stage, selection for

quality in done. If necessary, separate disease nurseries may be planted to evaluate disease resistance of the clone s.

Fourth to eighth years : Replicated yield trials are conducted at several locations along with suitable check. The yielding ability, quality and disease resistance etc. of the clones are rigidly evaluated. The best clones that are superior to the check in one or more characteristics are identified for release as varieties.

Ninth year : The superior clones are multiplied and released as varieties.

Advantages :

a. Varieties are stable and easy to maintain

b. Avoids inbreeding depression

c. Clonal selection, combined with hybridization generates necessary variability for several selections.

d. Only method to improve clonal crops

e. Hybrid vigour is easily utilized selection may be used in maintaining the purity of clones.

Disadvantages

a. Selection utilizes the natural variability already present in the population.

b. Sexual reproduction is necessary for creation of variability through hybridization.

c. Applicable only to the vegetatively propagative crops.

Problems in Breeding asexually propagated crops

1. Reduced flowering and fertility
2. Difficulties in genetic analysis
3. Perennial life cycle.

Genetic variation within a clone

Genetic variation within a clone may arise due to :

1. Mutation
2. Mechanical mixture
3. Sexual Reproduction 115

1. **Mutation :** The frequency is generally very low (10-5 to 10-7). Ordinarily dominant mutations would be expressed in the somatic tissue. A mutant allele would be homozygous only when.

i) both the alleles in a cell mutate at the same time producing the same mutant allele or

ii) The mutant allele is already in heterozygous condition in the original clone. Though rare, both these events are possible. Bud mutations may aften produce chimeras i.e. an individual containing cells of two or more genotypes. But mutations make possible selection of buds to establish new desirable clones, the process being known as Bud selection. It is of some importance in improvement of perennial crops like fruit trees or of those crops where flowering does not take place. It requires large number of plants to be observed and several trained persons to detect the mutant buds. Hence the bud selections are practiced in commercial plantations.

2. **Mechanical mixtures :** Mechanical mixtures produces genetic variation within a clone much in the same way as in the case of purelines.

3. **Sexual reproduction :** Occasional sexual reproduction would lead to segregation and recombination. The seedlings obtained from sexual reproduction would be genotypically different from the asexual progeny. It is evident that only clones would tend to became variable atleast in annuals and biennials. Eg. Potato

Clonal degeneration : The loss in vigour and productivity of clones with time is known as clonal degeneration and results due to :

1. Mutation
2. Viral diseases
3. Bacterial diseases

Achievements

I. Through clonal selection :

Potato : 1.Kufri Red from Darjeeling Red Round

2. Kufri Safed from phulwa

3. Bombay Green banana is a bud selection from dwarf Cavendish :

Pidi Monthan from Monthan

II. Through hybridization :

Potato : Kufri Alankar, Kufri Kuber, Kufri Sindhuri, Kufri Kundan, Kufri Chamatkar, Kufri Jyothi (late blight resistant), Kufri Sheetman (frost resistant)

Sugarcane : Co 1148, Co 1158, CoS 510, Co 975, Cos 109, Co 541

Mango : Pedda Neelam, Chinna Suwarnarekha

Banana : High gate from Gross Michel

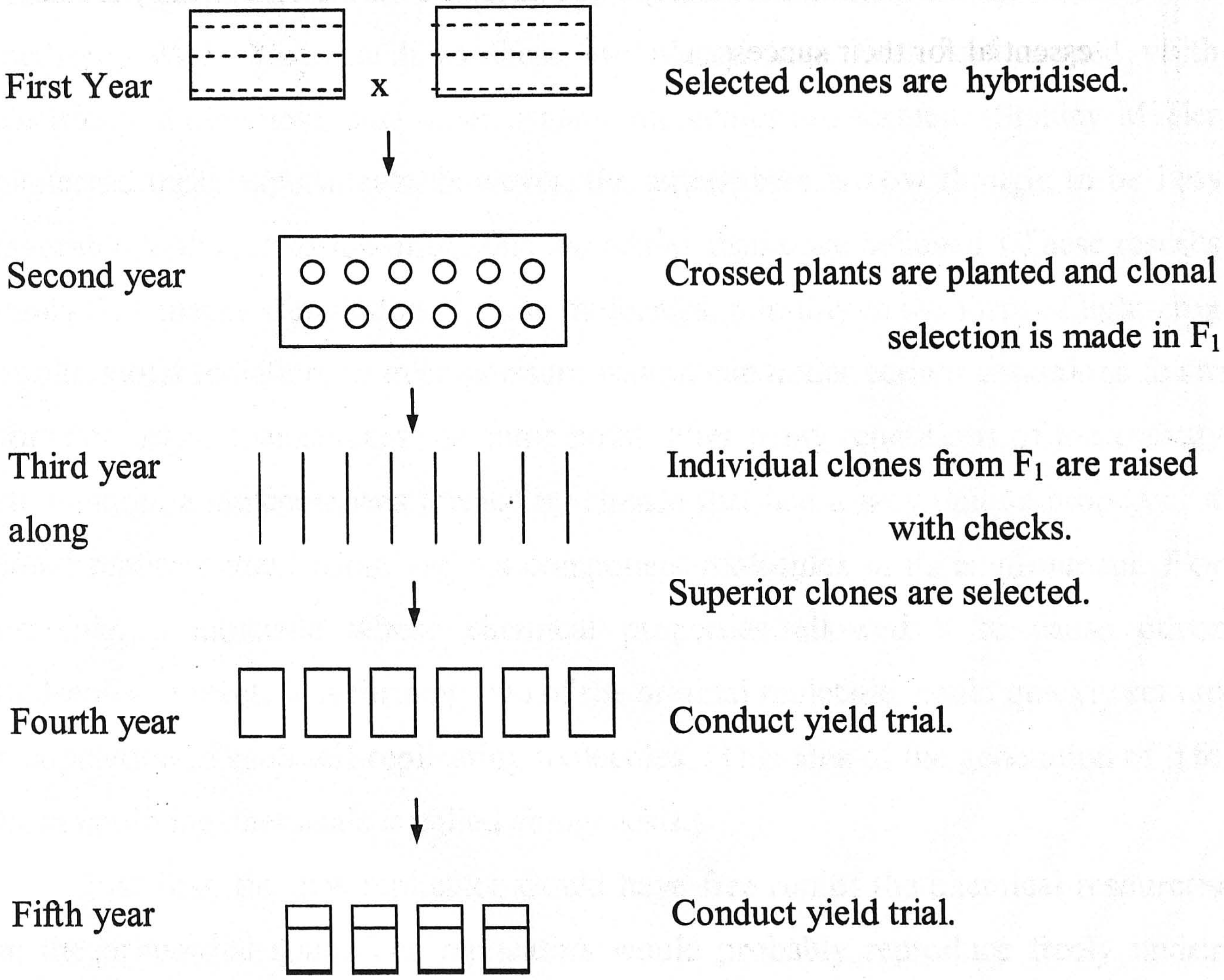
Citrus : Robertson Navel Orange

Sweet oranges : Yuvaraj blood Red

Turmeric : Kesari, Kasturi

Hybridization and selection

a) Inter varietal or Intra Specific Hybridization



b) Inter Specific hybridization

Successfully used in many crops like Rubus, Malus Strawberries. Potato variety Kufri Kuber was developed from a complex cross. (S. *Cuttilobum*) x (S.*tuberosum*) x S.*andigena* Generally interspecific crosses are useful for disease resistance. For E.g. S.*demissum* has been extensively used to induce late blight resistance in potato.

Sugar cane : All varieties of sugar cane now in cultivation have been developed from complex cross between S.*officinarum*, S.*barberi*, S. *robustum* and S.*spontaneum*. The success of interspecific hybridization in clonal crops are due to following reasons.

a. Due to asexual: reproduction there is no segregation.
b. Most of them are not seed crops and hence flowering and fertility are not essential for their success.

UNIT V: ORIGIN OF LIFE

Introduction

Many theories have been advanced as explanations for the origin and history of life on Earth. The most popular with the majority of evolutionary theorists is the primordial soup theory, which states that self-replicating entities, the precursors to life as we know it, arose spontaneously out of the chemical environment of the early Earth. This theory argues that the chance reactions taking place at high rates in the chemical mixture of the early atmosphere eventually gave rise to molecules with the property of replication.

The Primordial Soup Theory

The exact environment of the early atmosphere on Earth is not known, but some chemicals have been hypothesized: carbon dioxide, ammonia, water, and methane. When gases such as these are placed in a flask and zapped with electricity, amino acids and other organic molecules are formed. (Stanley Miller pioneered these experiments; however, the atmosphere is now thought to be less favorable to the formation of organic molecules than once believed.) These results imply that energy stimulation of these molecules, possibly in the form of lightning or ultraviolet radiation, or even pressure waves, can under certain conditions form complex organic substances. At some point, after many repetitions of the energy stimulation, a molecule was formed by chance that had a very unique property: it could replicate itself using various component molecules in its environment. For example, a molecule whose chemical properties allowed it to cause other molecules to react, thus forming two of the original molecule, could quickly set up a population of such self-replicating molecules. (This idea of the generation of life from nonliving chemicals is called abiogenesis.)

At first, the new replicator would have free run of the chemical resources in the primordial soup. The replicators would probably reproduce freely under such conditions because they would have a monopoly on the available resources. In addition, the sheer numbers of copies being made, added to the fact that the replicator would be very primitive and without editing mechanisms, would result

in numerous copying errors. These errors are mutations that will later be used in the development of natural selection.

After a while, the replicators would have used up many of the available chemicals in the soup (especially in localized areas), and they came into direct competition with each other for the use of the remaining resources. This set up a situation where natural selection could begin to operate on the population. Those replicators that were better at reproducing, either because they reproduced faster, out-competed others for resources, used fewer chemicals, etc., came to dominate the soup. Natural selection favored those that possessed traits allowing them to out-replicate the others.

After a long time, the resources of the area would become extremely scarce. Selection pressure would come into existence strongly favoring those replicators that could synthesize at least some of their own chemicals needed for reproduction. Once a certain chemical became scarce, selection pressure would favor those replicators that produced the chemical themselves.

Eventually, it seems reasonable to theorize that replicators might begin manufacturing enough chemicals that they would want to keep all the chemicals close by. One way to do this is to manufacture a container (made of chemicals that the replicator can already manufacture). This was the precursor to modern cells.

Other replicators would have evolved alternative strategies for dealing with the intense competition for resources. They may have developed ways of manufacturing chemicals that broke down other replicators into their component structures, thus providing raw materials from the destruction of competitors. Still others may have evolved new chemicals with new and improved properties and thus out-competed the other replicators.

Co-operative Replicators and the Development of Vehicles

Over time, replicators probably began to cooperate among themselves, with multiple replicators existing together in an aggregate, each producing a different chemical or performing a different function. These "replicator teams" quickly came to incorporate the container idea mentioned above, for the dual reason of keeping chemicals in one area and holding the replicator team together.

In time, under the influence of continual competition and selection, the replicators would come to develop more and more organized, advanced structures. These structures would most likely be devoted to manufacture of chemicals, storage, repair, etc. Cell organelles were born.

With the birth of the proto-cell, the first vehicle would arise. With organelles, replicators would become more and more specialized while developing greater and greater complexity in their biochemical pathways. The end result of all this developmental activity is the ancestral bacterium.

The Development of Eukaryotes

Bacteria, the most primitive of modern-day cells, lack a nucleus to direct and regulate cell function and are thus called prokaryotes. This is consistent with the model of their evolution: various replicators coming together and cooperating, but without the guidance of a central system. However, to form eukaryotic cells (cells with nuclei), a guidance system is needed.

This guidance system will most likely arise from the replicators themselves eventually. The replicators, increasingly specialized, will be too numerous in a complex cell to simply float about, and they will concentrate in one region. This region may develop a membrane for protection, and the nucleus is born.

Further development of the eukaryotic cell will take place when some of them begin to ingest other cells as sources of chemicals. At some point, the other cells might not have been broken down by the proto-eukaryote's digestive chemicals. These ingested cells might then provide extra energy to the larger cell, thus establishing an endosymbiotic relationship. This is believed to have taken place in the incorporation of mitochondria, chloroplasts, and even cilia into eukaryotic cells.

Theories of Evolution

Evolution is the change in the inherited characteristics of biological populations over successive generations. Charles Darwin and Alfred Wallace were the first to formulate a scientific argument for the theory of evolution by means of

natural selection. Evolution by natural selection is a process that is inferred from three facts about populations:

1) more offspring are produced than can possibly survive,

2) traits vary among individuals, leading to different rates of survival and reproduction, and

3) trait differences are heritable.

Thus, when members of a population die they are replaced by the progeny of parents that were better adapted to survive and reproduce in the environment in which natural selection took place. This process creates and preserves traits that are seemingly fitted for the functional roles they perform. Natural selection is the only known cause of adaptation, but not the only known cause of evolution. Other, nonadaptive causes of evolution include mutation and genetic drift.

The occurrence of evolution is undeniably established by the various types of evidences that are available today. However, how it has come about is explained by the several ideas that have been put forth from time to time. These ideas are known as 'theories on (organic) evolution'. Following are the main theories on the mode of evolution.

Lamarckism

It is the name given to the theory proposed by Jean Baptiste Lamarck (1744-1829), the French Zoologist. It is also known as the 'theory of inheritance of acquired characters'. Lamarck has explained his ideas on organic evolution in the book Philosophica Zoologique, published in 1809. He recognised the fundamental continuity underlying the diversity in animals. His theory is the result of his systematic studies and is based on the following three main ideas.

New Needs

Lamarck suggested that new needs arise in the organisms as the ever-changing environment influences them. Changes in the environment create a new need in the organism for adapting to that change, failing which the organism cannot survive.

Use and Disuse of Organs

The appearance of new needs forces the organism to put in additional efforts to fulfil the need. It may become necessary for the organism to put a particular part of its body into more and more frequent use or less and less frequent use. This idea came to be known as use and disuse of organs. Lamarck believed that any particular part of the body that is put into more frequent use, has a tendency to develop stronger and better while any part of the body, which is put into less or no use, has a tendency to gradually disappear. Thus, by a differential use and disuse of various body parts during its life span, an organism would change to some extent and acquire some new characters.

Inheritance of Acquired Characters

According to Lamarck the characters acquired by an organism are transmitted by heredity to the next generation. In every generation, fresh characters are acquired. With the result, after many generations, the changes accumulate to the extent that the species becomes modified into a new one.

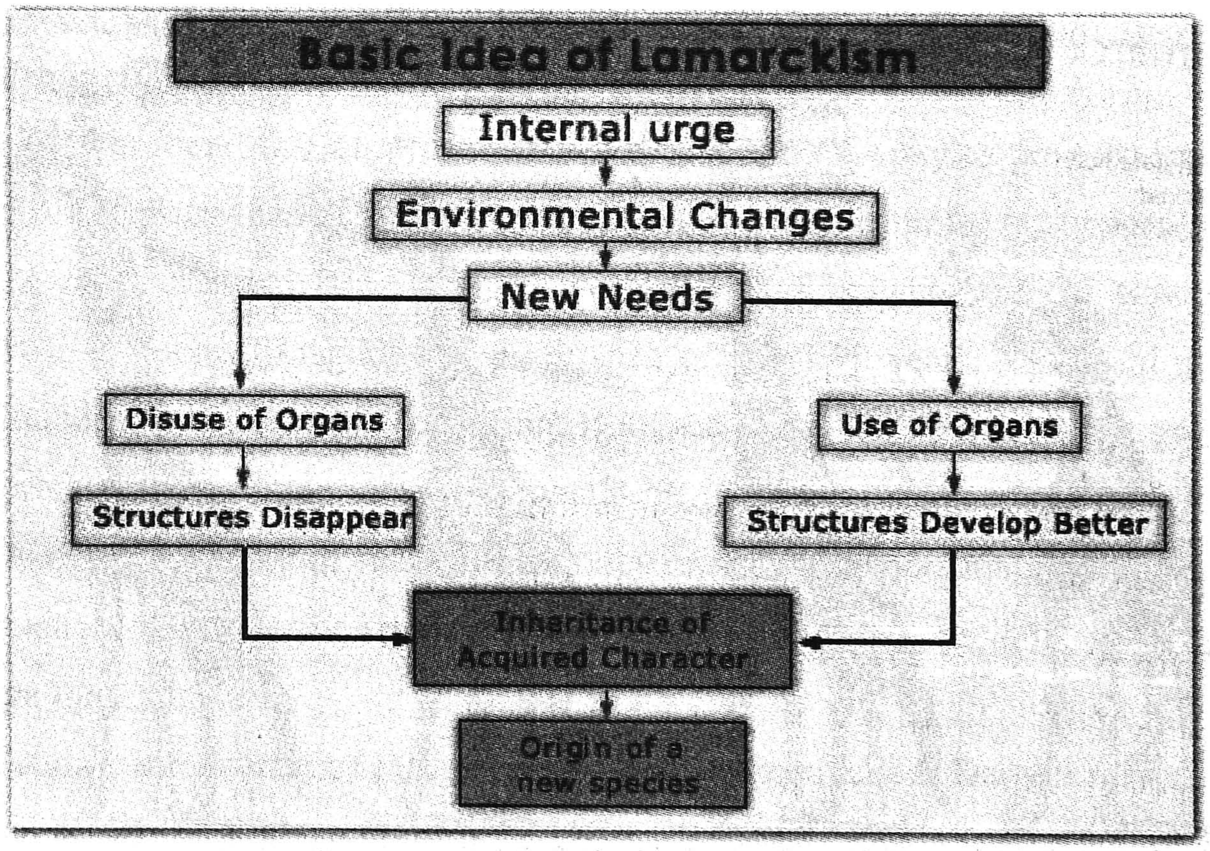
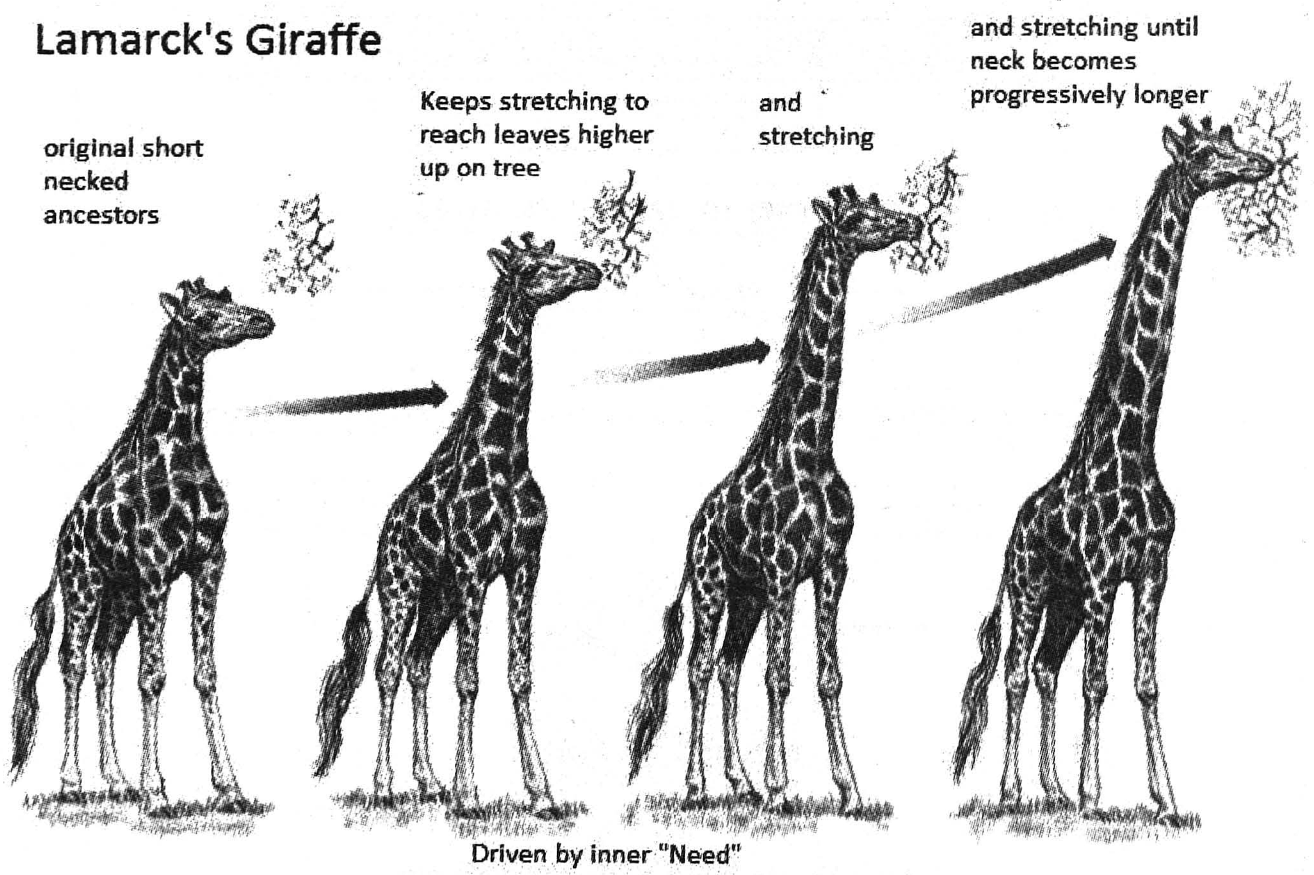


Fig: Basic idea of Lamarckism

Lamarck's ideas on evolution may be explained by citing a few examples from his own writings. The most familiar among them is of Giraffe. The Giraffe

differs from other mammals in having a long neck and longer fore limbs. The long neck and longer for limbs, according Lamarck, are the characters that Giraffe acquired in its lifetime. Lamarck suggested that Giraffes started stretching their neck and fore limbs in order to reach leaves of taller plants. They were forced to do so due to the competition that they were facing from other herbivores. The practice of stretching the neck and fore limbs continued for generation after generation resulting in a gradual increase in the length of the neck and fore limbs. This example is cited by Lamarck in support of his view that any organ put into continuous use has a tendency to develop better. Similarly, Lamarck interprets the absence of limbs in the snakes as an evolutionary change. He suggested that the ancestors of snakes, which had well-developed limbs, faced the problem of increased predation. In order to protect themselves, snakes started crawling on the ground into crevices and holes. Continuous disuse of limbs resulted in limbs becoming shorter and shorter and finally they disappeared in one generation.



Criticism of Lamarckism

August Weismann a German biologist tested the Lamarck's theory by cutting off the tails of the mice generation after generation. If Lamarck's theory

was correct, the subsequent generations should have developed shorter and shorter tails. No such shortening of tails was observed. However, even after a very large number of generations, mice continued to develop tails of the same length.

Darwinism

It is the name given to the theory proposed by Charles Darwin (1809-1882). It is also known as the theory of natural selection. He formulated this theory along with another English biologist, Alfred Russel Wallace (1823-1913) in the year 1858. This theory is the result of enormous amount of natural history collected by Charles Darwin on different species of plants and animals, during his voyage on the ship - H.M.S Beagle which lasted for about 5 years.

The theory was greatly influenced by the essay on population by Thomas Malthus, which underlined the competition between species for obtaining food. Darwin has given a vivid explanation of his ideas on evolution in the book 'On the Origin of Species by means of Natural Selection' published in 1859. The theory of natural selection is based on the following principles.

Over Production

Darwin found that all organisms have a natural capacity to produce more number of offspring than that survive. He quoted several examples to support this idea, the most classical being that of the elephant. Elephants are the slowest breeders in the animal kingdom. In a life span of about 90 years, a parental pair may produce a maximum of 6 offspring. If all the young of a pair survive, grow and reproduce at the same rate, there would be 19 million elephants in just 750 years! However, in spite of such an enormous reproductive potential, the number of individuals in each species remains nearly constant over long periods of time.

Struggle for Existence

As a consequence of this prodigal rate of reproduction, Darwin reasoned that there must be a competition between individuals belonging to the same or other species, for sharing food, water and living space. Every individual puts efforts for fulfilling the basic needs such as space for living, food for living, mate for reproduction and also protection from enemies. This competition is known as 'struggle for existence'.

Variations

The differences that an organism shows from its parent or from its related species are called variations. Variations are the rule of nature. Variations help an animal to adjust better to its environment. An animal, which has developed a favourable variation, has a better chance of survival and its offsprings are likely to inherit these variations.

Survival of the fittest

Only those organisms, which have favourable variations, are at a specific advantage over others. Such individuals survive the struggle for existence while those, which are less adapted or not adapted, get eliminated. This idea came to be known as 'survival of the fittest'.

Natural selection

Darwin believed that nature selects only those individuals, which have favourable variations and thereby have competitive advantage over others. This process is known as natural selection. The members of any particular species can develop and survive only when they are able to adapt themselves to the changed environmental conditions, by virtue of their favourable variations. Such variations are inherited and after a number of generations the variations become so prominent that a new species has been formed. Thus, new species develop from the existing ones in a slow and gradual way. This process of formation of a new species is called speciation. Darwin believed that while variations are the raw materials for evolution, natural selection is the force responsible.

Illustration of darwinism central theme

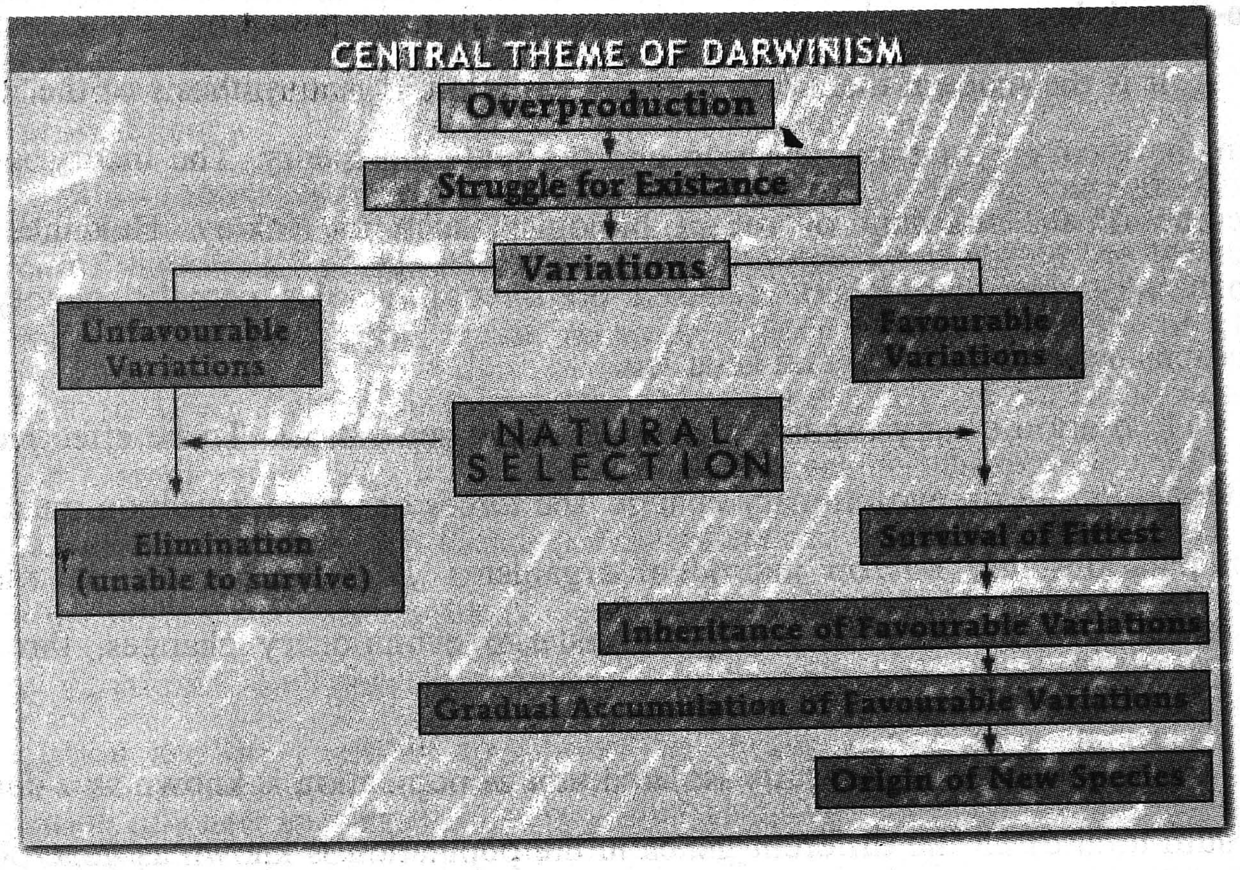


Fig. Central Theme of Drawinism

Darwin firmly believed that the environment would gradually weed out organisms, which have unfavourable variations, or with no variations, but preserve only those with favourable variations. However, Darwin failed to recognize the source of variations in the organisms. Hence goes the common saying - 'Darwin explained the survival of fittest but not their arrival'.

Drawbacks and criticism of Darwinism

Darwinism held that small variations form the raw material for evolution, this means that the organs appear as small structures and gradually become perfect after several generations.

In certain animals, some organs have developed beyond the stage of usefulness. Eg., Antlers of deer, tusks of elephants etc. These organs instead of providing usefulness to the possessors, offer hindrance in their daily life.

According to the theory of natural selection, only the useful organs are selected in the struggle for existence. The vestigial organs serve no function, yet they are preserved generation after generation.

Neo-darwinism

It is the modern theory on organic evolution. It is a synthesis of the ideas from the theory of natural selection and the mutation theory. The theory is the outcome of the thinking of recent biologists such as Huxley, Haldane and Dobzhansky. Neo-darwinism is based on three principles namely; genotypic variations, natural selection and isolation.

Evolution of any organism requires many generations. To be effective in producing long-term evolutionary change, the forces of natural selection must act on populations and not on individual organisms. It is the population, which possesses the variability necessary to undergo evolutionary changes, through space and time.

A more or less genetically isolated unit of population is known as a deme. The sum total of all the different genes in the population is known as gene pool. The evolutionary future of any population depends upon its gene pool.

The frequency of any given gene in a deme, relative to all its alleles in the same locus, is known as gene frequency. In spite of a continuous movement of genes, called gene flow caused by in-breeding, the gene frequencies tend to remain the same. This state of stability is known as genetic equilibrium. Under these circumstances the deme remains genetically unaltered and no evolution occurs in that deme. Evolution takes place only if and when the genetic equilibrium is altered as a result of mutations, environmental changes and natural selection. Two more factors may be responsible for altering the genetic equilibrium. Sometimes variations may occur in the gene frequency, particularly in small population. Such a variation is known as genetic drift. It can produce significant changes in small populations. Another factor, which may alter the genetic equilibrium, is called isolation. It involves the separation of the sub-groups of a population having different genotypes, from the population having different genotypes, from one another. Speciation can occur only if and when the population splits into two or more separate demes, each with its own gene pool. If the populations are isolated, mutation and natural selection can occur independently on them and each of them can develop into a new species.

Natural selection is responsible for maintaining constancy of a species, as well as in changing it. It maintains constancy as long as the environment remains stable, but promotes the formation of new species, if and when the environment changes.

Mutation Theory

Since Darwin's theory failed to explain certain facts, particularly the source of variations, several scientists started looking for bridging the lacuna in the theory. One such effort is the mutation theory proposed by Victor Hugo DeVries in 1902. According to this theory, new species arise from the existing one, in a single generation by a sudden change described as mutation. In plant, *Oenothera lamarckiana*, he found certain variations in one of the generations, from the normal plant. Such plants were found to breed truly giving rise to some more variations. DeVries suggested that large-scale variations occurring in a population are the result of mutations that occur frequently, resulting in the formation of a new species.

Natural selection is dependent on there being different variations of the same trait available. Additionally, these variations must confer different survival advantages or disadvantages. Natural selection could not operate on a uniform population that did not vary in any characteristic, because no organism would have a survival advantage. Similarly, if different characteristics existed but made absolutely no difference in survival and reproduction, natural selection could not operate because the possession of one characteristic or another would not confer a survival advantage.

Therefore, natural selection is dependent on the existence of mutations in the genes coding for different characteristics of an organism. Most mutations in DNA are spontaneous and random, sometimes caused by passing cosmic rays or other exposure to radiation. Mutations may also be caused by errors in the formation of the genes in the parents' gametes in sexual organisms. Additionally, "proofreading" enzymes built into many higher organisms sometimes fail, leaving an incorrect version of an organism's DNA. (Some lower organisms do not have a

proofreading step. Many viruses do not proofread copies of their genomes, resulting in their extremely rapid mutation rate.)

A vast majority of mutations in an organism's DNA have deleterious effects on the organism and thus will be immediately selected against, or they will be irrelevant or have only very marginal effects. Only a tiny percentage of all mutations will confer a survival advantage on the organism that inherits it. Even these mutations generally change very little about the organism's structure or function. A small change is much more likely to be beneficial (or at least not harmful) to an organism than a large random change in a system that already works even somewhat well. For a good comparison, think of a car or other complex machine. The vast majority of alterations you could make would be harmful (for example, removing the engine, smashing the windshield, etc.). Another percentage will be irrelevant (for example, scratching the paint, denting the door, etc.) Only a tiny fraction of the possible alterations would result in an improvement of the car's function (for example, adding antilock brakes).

There is another important distinction to be made between different types of mutations. The type of mutation most people think of when presented with the word "mutant" is called a macromutation, or a mutation that involves a very large change. A frog born with eyes in its throat or with extra legs would be an example of a macromutant. Macromutations are equivalent to taking a blind leap over a precipice and hoping to land on a ledge. They are virtually never beneficial (no well-documented case exists) and are not the forces of change in natural selection. The (erroneous) belief that macromutations drive evolutionary change is called saltationism, and is generally discredited today. The second type of mutation is called a micromutation, or a mutation that involves a very small change. An incredibly vast majority of all mutations fall into this category. These mutations can be (and generally are) harmful in effect, but are not drastic changes, but rather fine gradations. Micromutations are what evolutionists discuss when studying natural selection.

Natural selection is quick to seize upon the very rare beneficial mutations that arise. Even a small survival advantage will be selected for over generations,

eventually saturating the overall gene pool with the altered gene. Natural selection may depend on random mutations, and its operation may be slow and fitful, but it is extremely efficient in "weeding out" successful adaptations to be passed on to future generations.

Characteristics of Mutation Theory

De Vries's mutation theory has the following characteristics :

a. Mutations appear from time to time among the organisms of a naturally breeding species or populations. The organisms with mutations are called mutants. These mutants are clearly distinct from their parents.
b. Mutations are heritable and form new species. They do not disappear by crossing.
c. Mutations arise suddenly in one step, i.e., new species arise suddenly in one step and not gradually.
d. Mutations occur in all possible directions and may be advantageous or disadvantageous.
e. Unsuitable mutants are destroyed by natural selection.
f. Mutations appear full-fledged and, hence, there is no question of incipient stages in the development of an organ.

Advantages of Mutation Theory

The advantages of de Vries's mutation theory are the following :

a. The mutation theory describes the importance of mutation in selective value of organisms.
b. The mutation theory explains the occurrence of evolutionary changes within short period in contrast to natural selection (which describes slow and continuous variations).
c. Mutation theory explains the absence of connecting links as no criteria against evolution but its possibility exist.
d. Occurrence of mutations in large and divergent direction removes the possibility of species disappearance by crossing.
e. Since mutations appear fully formed from the beginning, there is no difficulty in explaining the incipient stages in the development of organs.

f. Mutation is of great service to breeders in developing new useful varieties.

Objections to Mutation Theory

a. The mutation theory was unable to explain the presence of flightless birds on oceanic islands.
b. It could not explain the existence of discontinuity in distribution among individuals.
c. Many mutations, described by de Vries in Oenothera lamarckiana, are now known to be due to certain numerical and structural changes in the chromosomes. For instance, "gigas" mutant of O. lamarckiana was later found to be due to polyploidy.
d. Mutation theory alone could not explain evolution. It, however, provided raw material for other forces to act upon it and bring about evolutionary changes.

Modern Synthetic Theory

The modern synthetic theory of evolution involves five basic processes — mutations, variations, heredity, natural selection and isolation. In addition, three accessory processes affect the working of these five basic processes. Migration of individuals from one population to another as well as hybridization between races or closely related species both increase the amount of genetic variability available to a population. The effects of chance acting on small populations, may alter the way in which natural selection guides the course of evolution (Stebbins, 1971).

1. Mutation. Alteration in the chemistry of gene (DNA) is able to change its phenotypic effect (i.e., nature of polypeptide) is called point mutation or gene mutation. Mutation can produce drastic changes or can remain insignificant. There are equal chances of a gene to mutate back to normal. Most of the mutations are harmful or deleterious and lethal but not all. Most of the mutant genes are recessive to normal gene and these are able to express phenotypically only in homozygous condition. Thus, point mutations tend to produce variations in the offspring.

2. Variation (Recombination). The nature of genetic variations caused by reshuffling of genes during sexual reproduction (recombination) was very little known at the time of Darwin. Recombination — that is, new genotypes from already existing genes — is of several kinds :

(a) the production of gene combinations containing in the same individual two different alleles of the same gene, or the production of heterozygous individuals (meiosis).

(b) the random mixing of chromosomes from two parents to produce a new individual (sexual reproduction)

(c) the mixing of a particular allele with a series of genes not previously associated with it, by an exchange between chromosomal pairs during meiosis, called crossing over, to produce new gene combinations. Chromosomal mutations such as polyploidy, deletion, duplication, inversion and translocation also result in variation.

3. Heredity. The transmission of characteristics or variations from parent to offspring, is an important mechanism of evolution. Organisms possessing hereditary characteristics that are helpful, either in the animal's native environment or in some other environment that is open to it, are favoured in the struggle for existence. As a result, the offsprings are able to benefit from the advantageous characteristics of their parents.

4. Natural selection. Natural selection brings about evolutionary change by favouring differential reproduction of genes. Differential reproduction of genes produces change in gene frequency from one generation to the next. Natural selection does not produce genetic change, but once genetic change has occurred it acts to encourage some genes over others. Further, natural selection creates new adaptive relations between population and environment, by favouring some gene combinations, rejecting others and constantly moulding and modifying the gene pool.

The workings of natural selection are exceedingly complex because of the range of organizational levels at which it functions. Selection discriminates among available reproducible biotic entities to produce more efficiently adapted units.

Natural selection operates upon every stage in the life history of an organism. It produces non-random differential reproduction of biological units and may affect any biotic entity from the molecular to the community level.

Examples of levels at which natural selection makes differential discrimination are the following : intermolecule, intergene, interchromosome, intergamete, interindividual (Darwinian selection), interdemic, interracial, interspecific and intercommunity. Darwinian selection may result from differential natality among others.

5. Isolation. Isolation of organisms of a species into several populations or groups under psychic, physiological or geographical factors is supposed to be one of the most significant factors responsible for evolution. Geographical isolation includes physical barriers such as high mountains, rivers, oceans and long distances preventing interbreeding between related organisms. Physiological barriers help in maintaining the individuality of the species, since these isolations do not allow the Content interbreeding amongst the organisms of different species. This is called reproductive isolation. Speciation (Origin of new species). The populations of a species present in the different environments and are segregated by geographical and physiological barriers, accumulate different genetic differences (variations) due to mutations (both point and chromosomal), recombination, hybridization, genetic drifts and natural selection. These populations, thus, become different from each other morphologically and genetically, and they become reproductively isolated, forming new species.

The Different Types Of Evolution

1. Cosmic Evolution: The origin of time, space and matter, by the Big Bang
2. Chemical Evolution: The origin of higher elements from hydrogen.
3. Stellar and Planetary Evolution: The origin of stars and planets.
4. Organic Evolution: The origin of Life.
5. Macro-Evolution: The changing from one kind of species to another kind of species.
6. Micro-Evolution: The variation within kinds of species.

Biological evolution over time can follow several different patterns. Factors such as environment and predation pressures can have different effects on the ways in which species exposed to them evolve. Evolutionary biologists have labelled these differing patterns as divergent, convergent, and parallel evolution.

Divergent Evolution

When people hear the word "evolution," they most commonly think of divergent evolution, the evolutionary pattern in which two species gradually become increasingly different. Divergent evolution occurs when a group from a specific population develops into a new species. In order to adapt to various environmental conditions, the two groups develop into distinct species due to differences in the demands driven by the environmental circumstances. On a large scale, divergent evolution could be responsible for the creation of the current diversity of life on earth from the first living cells. On a smaller scale, it could responsible for the evolution of humans and apes from a common primate ancestor. On a molecular scale, it could be responsible for the evolution of new catalytic functions of enzymes and membrane protein topology.

Divergent evolution and speciation

If different selective pressures are placed on a particular organism, a wide variety of adaptive traits may result. If only one structure on the organism is considered, these changes can either add to the original function of the structure, or they can change it completely. Divergent evolution leads to speciation, or the development of a new species. Divergence can occur when looking at any group of related organisms. The differences are produced from the different selective pressures. Any genus of plants or animals can show divergent evolution. An example can involve the diversity of floral types in the orchids. The greater the number of differences present, the greater the divergence. Scientists speculate the greater that two similar species diverge indicates a longer length of time that the divergence originally took place.

Examples of divergent evolution

There are many examples of divergent evolution in nature. If a freely-interbreeding population on an island is separated by a barrier, such as the presence of a new river, then over time, the organisms may start to diverge. If the opposite ends of the island have different pressures acting upon it, this may result in divergent evolution. Or, if a certain group of birds in a population of other bird of the same species varies from their migratory track due to abnormal wind fluctuations, they may end up in new environment. If the food source is such that only birds of the population with a variant beak are able to feed, then this trait will evolve by virtue of its selective survival advantage. The same species in the original geographical location and having the original food source do not require this beak trait and will, therefore, evolve differently. Divergent evolution has also occurred in the red fox and the kit fox. While the kit fox lives in the desert where its coat helps disguise it from its predators, the red fox lives in forests, where the red coat blends into its surroundings. In the desert, the heat makes it difficult for animals to eliminate body heat. The ears of the kit fox have evolved to have greater surface area so that it can more efficiently remove excess body heat. Their different evolutionary fates are determined primarily on the different environmental conditions and adaptation requirements, not on genetic differences. If they were in the same environment, it is likely that they would evolve similarly. Divergent evolution is confirmed by DNA analysis where the species that diverged can be shown to be genetically similar.

The human foot evolved to be very different from a monkey's foot, despite their common primate ancestry. It is speculated that a new species (humans) developed because there was no longer was a need for swinging from trees. Upright walking on the ground required alterations in the foot for better speed and balance. These differing traits soon became characteristics that evolved to permit movement on the ground. Although humans and monkeys are genetically similar, their natural habitat required different physical traits to evolve for survival.

Convergent Evolution

Convergent evolution causes difficulties in fields of study such as comparative anatomy. Convergent evolution takes place when species of different ancestry begin to share analogous traits because of a shared environment or other selection pressure. Environmental circumstances that require similar developmental or structural alterations for the purposes of adaptation can lead to convergent evolution even though the species differ in descent. These adaptation similarities that arise as a result of the same selective pressures can be misleading to scientists studying the natural evolution of a species. Convergent evolution also creates problems for paleontologists using evolutionary patterns in taxonomy, or the categorization and classification of various organisms based on relatedness. It often leads to incorrect relationships and false evolutionary predictions.

Examples of convergent evolution

(1) Pterosaur

(2) Bat

(3) Bird

One of the best examples of convergent evolution involves how birds, bats, and pterosaurs (all different taxa that evolved along distinct lineages at

different times) came to be able to fly. Importantly, each species developed wings independently. These species did not evolve in order to prepare for future circumstances, but rather the development of flight was induced by selective pressure imposed by similar environmental conditions, even though they were at different points in time. The development potential of any species is not limitless, primarily due to inherent constraints in genetic capabilities. Only changes that are useful in terms of adaptation are preserved. Yet, changes in environmental conditions can lead toward less useful functional structures, such as the appendages that might have existed before wings. Another change in environmental conditions might result in alterations of the appendage to make it more useful, given the new conditions.

Parallel Evolution

Parallel evolution occurs when unrelated organisms develop the same characteristics or adaptive mechanisms due to the nature of their environmental conditions. Or stated differently, parallel evolution occurs when similar environments produce similar adaptations. The morphologies (or structural form) of two or more lineages evolve together in a similar manner in parallel evolution, rather than diverging or converging at a particular point in time.

Examples of parallel evolution

One example is the complex plumage patterns that seem to have evolved independently among many very different bird species.

A molecular example of Parallel Evolution is the ligand specificity of repressors and periplasmic sugar-binding proteins.

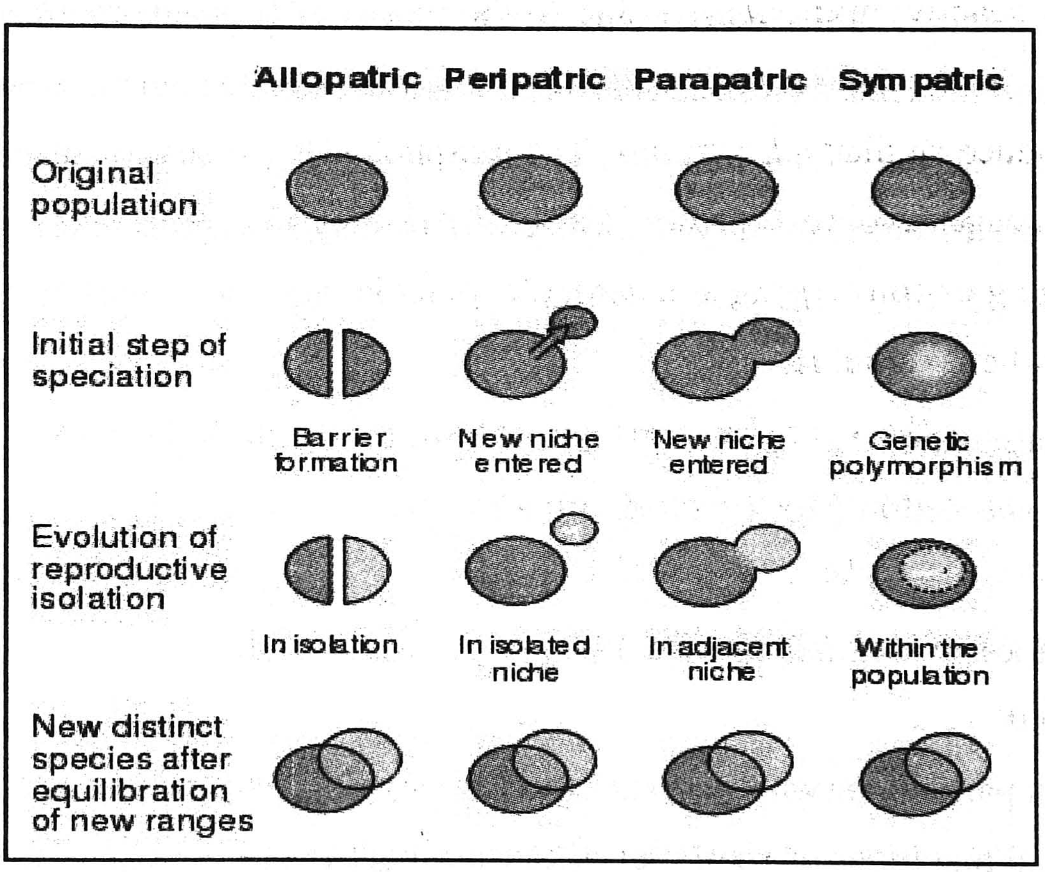
Types of speciation

Speciation can take place in two general ways. A single species may change over time into a new form that is different enough to be considered a new species. This process is known as anagenesis. More commonly, a species may become split into two groups that no longer share the same gene pool. This process is known as cladogenesis. There are several ways in which anagenesis and cladogenesis may take place. In all cases, reproductive isolation occurs.

Sympatric Speciation

Sympatric speciation occurs when populations of a species that share the same habitat become reproductively isolated from each other. This speciation phenomenon most commonly occurs through polyploidy, in which an offspring or group of offspring will be produced with twice the normal number of chromosomes. Where a normal individual has two copies of each chromosome (diploidy), these offspring may have four copies (tetraploidy). A tetraploid individual cannot mate with a diploid individual, creating reproductive isolation.

Sympatric speciation is rare. It occurs more often among plants than animals, since it is so much easier for plants to self-fertilize than it is for animals. A tetraploidy plant can fertilize itself and create offspring. For a tetraploidy animal to reproduce, it must find another animal of the same species but of opposite sex that has also randomly undergone polyploidy.



Allopatric Speciation

Allopatric speciation, the most common form of speciation, occurs when populations of a species become geographically isolated. When populations become separated, gene flow between them ceases. Over time, the populations may become genetically different in response to the natural selection imposed by their different environments. If the populations are relatively small, they may experience a founder effect: the populations may have contained different allelic frequencies when they were separated. Selection and genetic drift will act differently on these two different genetic backgrounds, creating genetic differences between the two new species.

Parapatric Speciation

Parapatric speciation is extremely rare. It occurs when populations are separated not by a geographical barrier, such as a body of water, but by an extreme change in habitat. While populations in these areas may interbreed, they often develop distinct characteristics and lifestyles. Reproductive isolation in these cases is not geographic but rather temporal or behavioural. For example, plants that live on boundaries between very distinct climates may flower at different times in response to their different environments, making them unable to interbreed.

Mechanism of Speciation

A. Instantaneous speciation (through individuals)

1. Genetically

(a) By single mutation in asexual "species"

(b) By macrogenesis

2. Cytologically, in partially or wholly sexual species

(a) By chromosomal mutations or aberrations (translocation, etc.)

(b) By polyploidy

B. Gradual speciation (through populations)

1. Geographical speciation

2. Sympatric speciation

A. INSTANTANEOUS SPECIATION

The process of instantaneous speciation may be defined as the production of a single individual (or the offspring of a single mating) that is reproductively isolated from the species to which the parental stock belongs and that is reproductively and ecologically capable of establishing a new species population. Certain phenomena among rotifers, cladocerans and nematodes suggest the occasional occurrence of asexual speciation by mutation. For example, in the rotifer order Bdelloidea, there occurs no male individual in entire order and so it is assumed that the ancestral species of this order was parthenogenetic. This ancestral species evolved into over 200 species, about 20 genera and 4 families only due to asexual speciation by mutation.

Vegetative reproduction by growth and splitting, thus, might offer a favourable condition for instantaneous speciation. This form of reproduction, however, occurs only in some of the lowest groups of animals, such as sponges, coelenterates (hydroids), turbellarians and bryozoans, groups in which the taxonomy is as yet too uncertain for a study of speciation.

Instantaneous Speciation through Macrogenesis

The sudden origin of new species, new higher taxa, or quite generally of new types by some sort of saltation (a change by leap across a discontinuity) has been termed macrogenesis. The macrogenesis theory of speciation has been supported by some geneticists such as Goldschimdt and Schindewolf. According to this theory, the production of a new type by a complete geneticreconstruction (macrogenesis) or by a major "systemic mutation" or macromutation is the crucial event in speciation. Such an event will produce a "hopeful monster" (as Goldschmidt, 1940 called it) which will become the ancestor of a new evolutionary lineage. The believers of macrogenesis theory also proposed that all new types appear in the fossil record suddenly and abruptly. These types are not connected with the ancestral types by intermediates, they claim, and cannot be derived from them by gradual evolution.

Instantaneous Speciation through Chromosomal Aberrations

Closely related species often differ more conspicuously in their karyotype than in their morphology. Among aspects of the karyotype that differ are chromosome number; the number of metacentric or acrocentric chromosomes; the presence and kind of paracentric or pericentric inversions; or of supernumerary chromosomes, and just about every aspect of chromosomal aberrations.

Visualizing this fact most cytogeneticists believed that chromosomal mutations have a significant role in the causing of instantaneous speciation. This belief is based on two assumptions :

(1) that the degree of difference displayed by two species requires a speciation process of such drastic dimensions that only chromosomal mutation can qualify, and

(2) that reproductive isolation between two species cannot be achieved without chromosomal reorganization. Now, it is known that chromosomal rearrangements in most cases does not produce new species, however, only in few cases it produces speciation.

(i) Chromosomal rearrangement without speciation

Except deleterious chromosomal mutations, most kinds of chromosomal aberrations (mutations) lead to chromosomal polymorphism rather than to the development of isolating mechanisms. The paracentric inversions of Drosophila are a well-known example of such chromosomal polymorphism. Yet each species has its own species specific polymorphism, and only very rarely do even the most closely related species share the same chromosomal polymorphism. This fact underlines the drastic nature of the chromosomal aberration during much of speciation. This phenomenon suggests that there is no necessary correlation between chromosomal mutations and speciation, since either can occur without the other.

(ii) Speciation coinciding with a chromosomal mutation

If the chromosomal mutations during speciation would not have some selective (evolutionary) advantage, they would not have occurred so frequently in nature. Consequently, they have the following two advantages in chromosomal

speciation— (a) chromosomal mutations have the potential to serve as (or contribute to) isolating mechanisms, and (b) the locking up and protection of a particularly favourable gene complement through a chromosomal mutation may create a new supergene as Wallace, first of all recognized it. Both of these components of chromosomal speciation can subsequently be improved by natural selection, either during a period of segregation in a geographic isolate or during subsequent parapatric speciation or by both processes. The term parapatric stands for the population or species which remain geographically in contact but not overlapping and rarely or never interbreeding.

(a) Chromosomal mutations as potential new isolating mechanisms

Any change in the structure of the chromosomes is called chromosomal mutations or chromosomal aberrations, whether it is an inversion, translocation, duplication, or any other change in the linear sequence of the genes or in the mechanics of the chromosome (for instance, spindle attachment).

4. Instantaneous Specification through Polyploidy

Polyploidy is a multiplication of the normal chromosome number. It is very widespread among plants and is an important mechanism of speciation in the plant kingdom. Except conifers, one-third of all species of plants have arisen by polyploidy. Among animals the polyploidy is much rarer. Among animals it occurs only in those groups which reproduce parthenogenetically. In few parthenogenetically reproducing animal groups such as lumbricid earthworms, turbellarians and in certain groups of weevils polyploidy is the principal method of speciation.

B. GRADUAL SPECIATION

Gradual speciation is the gradual divergence of populations until they have reached the levels of specific distinctness. Two modes of gradual speciation have been postulated;

(a) one involving geographical separation of the diverging populations (geographic speciation), and

(b) the other involving divergence without geographic separation (sympatric speciation).

1. Geographic or Allopatric Speciation

Geographic or allopatric speciaton states that in sexually reproducing animals a new species develops when a population that is geographically isolated from the other populations of its parental species acquires during this periods of isolation characters that promote or guarantee reproductive isolation after the external barriers break down. The geographic speciation is the almost exclusive mode of speciation among animals, and most likely the prevailing mode even in plants, is now quite generally accepted.

2. Sympatric Speciation

The majority of authors until fairly recently considered sympatric speciation, that is, speciation without geographic isolation to be the prevailing mode of speciation. Such a speciation is based on two postulates :

(a) the establishment of new populations of a species in different ecological niches within
the normal cruising range of individuals of the parental population, and

(b) the reproductive isolation of the founders of the new population from individuals of the parental population. Gene flow between daughter and parental population is postulated to be inhibited by intrinsic rather than extrinsic factors.

Adaptation

Adaptation can be defined as occurrence of genetic changes in a population or species as a result of natural selection so that it adjusts to new or altered environmental conditions. Thus, any characteristic that is advantageous to a particular organism or population is called an adaptation.

ADAPTIVE RADIATION

The term adaptive radiation has been coined by H.F. Osborn (1902) for explaining evolution, from a single ancestor, of a number of descendants with a great variety of adaptations to different niches. More aptly stated the phenomenon of adaptive radiation is the diversification of a dominant evolutionary group into a large number of subsidiary types adapted to more restrictive modes of life (different adaptive zones) within the range of the larger group. According to

George Gaylord Simpson (1940, 1953), adaptive radiation is the rapid proliferation of new taxa (species) from a single ancestral group.

Polymorphism

One of the most characteristic features of any natural population is its diversity. This diversity is obvious when we consider the human species, for we are attuned to sensing differences in human appearance, personality, sexuality, and so on. In the populations of flies or dandelions such well marked diversity does not occur but it exists nevertheless. In genetic terminology natural populations are said to be polymorphic.

Polymorphism is most apparent when it affects a visible or behavioral phenotype, but is not at all restricted to such traits. R. Lewontin and J. Hubby, in 1966, undertook the first extensive analysis of protein polymorphisms in natural population of Drosophila pseudoobscura by subjecting extracts of individual flies to get electrophoresis and observing the rates of migration of various proteins, which represented 18 gene loci. They found, quite unexpectedly, that many of the proteins existed in the population in the form of isoelectric variants, meaning that for a given type of protein some individuals possessed a fast-migrating species and others a slow-migrating species. Numerous subsequent studies of such diverse species as barley, wild oat, horse-shoe crab, mouse and man have all produced the same result : an abundance of protein polymorphism is found wherever it is sought. Protein polymorphism signals the existence of allelism, and it has been estimated that 20 to 50 per cent of all structural gene loci in a given species exist in two or more allelic forms in any given population. The polymorphism may arise in a population by the following three basic avenues—transient polymorphism, balanced polymorphism and random fixations of natural mutations.

Transient polymorphism:

Transient polymorphism is a by-product of directional natural selection. If we imagine that allele a1 has a selective advantage over a2, then with time a1 should proceed toward fixation at $p = 1$, and a2 should proceed toward elimination at $q = 0$. While this process is occurring, both a1 and a2 will be present in the gene

pool and a1/a2 heterozygotes will be present in the population. As the name implies, transient polymorphism represents a temporary situation. For example, during the course of industrial melanization, both dark and light moths would be expected to cohabit the Manchester trees for the interim, but the proportion of light moths would be seen to diminish with time as dark moths gradually predominated.

Balanced polymorphism:

Balanced polymorphism is also relatively permanent kind of equilibrium in which alleles a1 and a2 are present in the population at some steady-state frequencies. Balanced polymorphism is originated by disruptive or diversifying selection and heterosis.

Random fixation of natural fixation:

The random fixation of natural fixation method of origin of polymorphism is also called Neutral Mutation–Random Genetic Drift hypothesis or "Non-Darwinian Evolution" and this idea has been developed by S. Wright and Kimura. This hypothesis is based on the following two assumptions : (1) The first assumption states that selectively neutral mutations can occur in genes that code for proteins. This will clearly be true in the case of "synonymous" mutations in which one codon is replaced by another codon dictating the same amino acids, but it is also proposed to be true in the case of mutations that lead to amino acid substitutions. The idea is that an acidic amino acid might occur in an "unimportant" region of the protein, with the result that the emergent mutant protein is identical to the original in all functional aspects. (ii) The second assumption states that neutral alleles, being neither selectively advantageous nor disadvantageous, simply drift in the gene pool. Thus, if a neutral mutation arises in woman's germ cell and this germ cell gives rise to a female child, the probability is about 0.5 that the mutant allele will be transmitted to a grandchild and 0.5 that it will not. If it is not, then q becomes equal to zero and the allele is lost. Polymorphism of cytochrome c of all eukaryotes and haemoglobin protein of all vertebrates strongly support the Wright-Kimura hypothesis.

Coevolution

Evolutionary changes in two or more interacting species are called coevolution. When organisms that are ecologically intimate -- for example, predators and prey, or hosts and parasites -- influence each other's evolution, we say that coevolution is occurring. Birds are often important actors in coevolutionary systems. For example, predation by birds largely drives the coevolution of model and mimetic butterflies. Some butterflies have evolved the ability to store poisonous chemicals from the food plants they eat as caterpillars, thus becoming distasteful. This reduces their chances of being eaten, since birds, once they have tried to devour such butterflies, will avoid attacking them in the future. Other butterflies have gradually evolved color patterns that mimic those of the distasteful butterflies (called "models"). It is disadvantageous for the models to be mimicked, because if the mimics become common then most of the butterflies with the model's color pattern taste good, the birds may resume attacking the models. Being tasted and spit out by a bird is a most dangerous experience for a butterfly. Therefore, mimicry presumably leads to a coevolutionary race -- the mimics evolving toward the color patterns of the models, and the models evolving away from the converging mimics. The birds actually may be directly involved in the entire coevolutionary complex, since they may be under selection for better powers of discrimination. Individuals that can tell the mimetic butterflies from the models will gain more nourishment at less cost in time and effort. Birds, of course, are presumed to be directly involved in many coevolutionary relationships with their competitors, predators, prey, and parasites. The relationship of seed-hoarding Clark's Nutcrackers and Pinyon jays with pinyon pines is a relatively well-studied example; and the evolution of long bills and sickle-shaped bills in some Latin American hummingbirds which match the long or sharply curved flowers from which they sip nectar (and which they pollinate) is another obvious case of coevolution. Hermit hummingbirds and the curved flowers of the genus Heliconia (seen increasingly as horticultural cut flowers) provide widespread and conspicuous examples of the latter phenomenon throughout the lowland moist forests of Central and South America.

Many fruit-eating birds, especially in tropical rain forests are coevolving with the plants whose fruits they eat. The birds get nourishment, and in the process the plants get their digestion-resistant seeds dispersed by regurgitation or along with the birds' droppings. Many characteristics of the plants have evolved to facilitate dispersal, and the behavior and diets of the birds have responded to those changes. In particular, the plants have evolved conspicuously colored, relatively odorless fleshy fruits to attract the avian dispersers of their seeds. They are coevolving in response to the finely honed visual systems of the birds; plant species coevolving with color-blind mammalian seed-dispersers have, in contrast, dull-colored but smelly fruits. The bird-dispersed plants often have evolved fruits with giant seeds covered by a thin, highly nutritious layer of flesh. This forces the bird to swallow the fruit whole, since it is difficult or impossible just to nip off the flesh. In response, birds that are specialized frugivores (that is, that do not take other kinds of food) have evolved both bills with wide gapes (so they can swallow the fruit whole) and digestive tracts that can rapidly dissolve the flesh from the large impervious seed, which then can be regurgitated. The most dramatic examples of avian coevolution are probably those involving brood parasites, such as cuckoos and cowbirds, and their hosts. The parasites have often evolved eggs that closely mimic those of the host, and young with characteristics that encourage the hosts to feed them. In response, some hosts have developed the ability to discriminate between their own and parasitic eggs, and various methods of destroying the latter. As one might expect, Brown-headed Cowbirds have their most serious impact on hosts, such as Kirtland's Warblers, that are thought to have only recently been subjected to cowbird attack and have not yet had time to evolve defensive reactions. Many examples of coevolution in response to competition between bird species can be inferred from studies of dietary habits and bill structures in various guilds of birds. Here, as in the other cases mentioned, direct evidence of coevolution is lacking. It is lacking for the same reason that there are very few cases of plain old single-population evolution actually being observed in nature. The process occurs over hundreds or thousands of generations, and extraordinary circumstances are required for it to be "caught in the act."